Anonos®

SPEED TO INSIGHT™

Lawfully & Ethically

A Blueprint for Harmonising Data Use
and Protection Under the GDPR

3rd Edition

www.SpeedToInsight.com

ABOUT THE AUTHORS

GARY LAFEVER

Gary LaFever is Co-Founder, Chief Executive Officer and General Counsel at Anonos, a former partner at the international law firm of Hogan Lovells and former Management Information Consultant at Accenture. Gary's 35+ years of technical and legal expertise enables him to approach data protection and utility issues from both perspectives. He is a co-inventor of 17 granted patents with over 70 additional patents pending in the U.S. and internationally.

MAGALI FEYS

Magali Feys is Chief Strategist of Ethical Data Use at Anonos and founder of AContrario Law, a boutique law firm specializing in IP, IT, Data Protection and Cybersecurity. In addition, Magali acts as a legal advisor of the Belgian Ministry of Health where she advises on privacy related matters and is a member of the legal working party e-Health of the Belgian Minister for Public Healthcare.

MARK LITTLE

Mark Little is Chief Data Strategist and Head of Engineering at Anonos, former Managing Director and Research Leader at CEB, now Gartner, and former Vice President, Operations at Digital Optics Corporation, now Broadcom.

v19_71220H

CONTENTS

FOREWORDS

The following forewords – appearing in the First Edition (January 2018 – Pre GDPR), Second Edition (January 2019 – Post GDPR) and Third Edition (June 2020 – 2 Years After GDPR) of this blueprint for harmonising data use and protection under the GDPR (this "Blueprint") – highlight:

- The importance of accountability-based information policy management.
- The need for new technologies like Anonos' first-of-its-kind patented BigPrivacy solution to support proportional use of data, responsive to the variety and complexity of different data uses.
- The minimum bar necessary to support unlocking data value while respecting the rights of individuals is the use of proven techniques and processes like Anonos BigPrivacy.

DATA PROTECTION MEGATRENDS

By Martin Abrams, Executive Director and Chief Strategist,
The Information Accountability Foundation (IAF) https://informationaccountability.org/

There are two data protection megatrends going on today. The first is the breaking wave of transformational data processing laws, regulations, and guidance evolving around the globe, epitomized by the GDPR. The second is the evolution of a data trust deficit into a full-fledged legitimacy conundrum. Yet people expect all the value of a highly observational world. How do global organisations reconcile the growing importance of data analytics, artificial intelligence, and machine learning with the increasingly complex and multi-jurisdictional regulations on lawful data use? And furthermore, how do they maintain trust that is based on both value and protection?

The Information Accountability Foundation believes accountability-based information policy management – being a trusted data steward – is a key element of the answer.

The GDPR requires accountability specifically. It requires organisations to have policies, and the processes to put those policies into effect. Those processes rest on new technologies that are demonstrable to assure conditions set by policy actually are actionable. The GDPR introduces these new controls in the form of technical and organisational measures necessary to support data protection by design and by default. Comprehensive data protection impact assessments that balance the interests of all stakeholders are a part of organisational controls.

Pseudonymisation, as newly defined under the GDPR, is another methodology that enables the fine-grained, risk-managed, use-case-specific controls necessary to support data protection by design and by default, particularly the fundamental data protection law principle of data minimisation. Data protection by design and by default embodies the goal of making technology controls that support appropriate uses.

A central core of data protection accountability and ethics is the will and ability to demonstrate that you can, in fact, keep your promises. Technologies that enforce data protection by design and by default show data subjects that in addition to coming up with new ways to derive value from data, organisations are pursuing equally innovative technical approaches to protecting data privacy – an especially sensitive and topical issue given the epidemic of data security breaches around the globe.

Vibrant and growing areas of economic activity – the “trust economy,” life sciences research, personalized medicine/education, the Internet of Things, personalization of goods and services – are based on individuals trusting that their data is private, protected, and used only for appropriate purposes that bring them and society maximum value. This trust cannot be maintained using outdated approaches to data protection. We must embrace new approaches like data protection by design and by default to earn and maintain trust and more effectively serve businesses, researchers, healthcare providers, and anyone who relies on the integrity of data.

Traditional approaches to data processing often involve the use of static identifiers that enable the ability to infer – or single out or link to – a data subject. This is because static identifiers, when used across multiple data sets, enable the overlay of the data sets so data that is not identifiable by itself, when combined with other overlapping data, leads to reidentification of a data subject. Conversely, data protection by design and by default can leverage dynamically changing identifiers to probabilistically prevent the ability to infer identifying information pertaining to a data subject across multiple data sets or data combinations – all in a manner that is capable of supporting mathematic analysis, audit, and enforcement.

New technologies are being introduced to implement data protection by design and by default. Anonos’ first-of-its-kind patented BigPrivacy solution is one example that supports proportional use of data in a manner that is responsive to the variety and complexity of different potential uses of data. Specifically, BigPrivacy can reveal different levels and types of information to the same and/or different parties at different times, for different purposes, at different places – and with respect to each, only as necessary for each use of data. By ensuring that only the minimum information necessary for each appropriate purpose is processed by “dialing-up” or “dialing-down” the linkability (or identifiability) of data, BigPrivacy helps to support accountable, ethical, fair, and legal data use.

ETHICAL TOOLS FOR CONTROLLING DISCLOSURE

By Jules Polonetsky, Chief Executive Officer.
The Future of Privacy Forum (FPF) https://fpf.org/

Writing in The New Yorker about the work of sociologist Beryl Bellman, Malcolm Gladwell said, “A secret isn’t invalidated by its disclosure, it’s defined by its disclosure. What makes a secret a secret is simply the operating instructions that accompany its movement from one person to the next.”

Today’s world is awash in secrets captured and disclosed by data-driven products and services. With all the personal information collected by wearables, smart homes, social media, smart cars, and innumerable other data-centric offerings, few companies are truly promising individuals’ privacy. Rather they are committing to responsible use of the data and controlled disclosure. The massive volume, variety and velocity of data created and captured by the ever-increasing numbers of data-driven offerings highlights the need for technical tools that enable those personal information commitments.

Traditionally, de-identification has been a primary method for enabling access to and use of data while protecting individuals’ privacy. De-identification has even sometimes been viewed as a “silver bullet” enabling organisations to reap the benefits of data processing while avoiding operational risks and legal requirements. However, scientists have repeatedly demonstrated that purportedly de-identified data sets can be vulnerable to reidentification attacks, thereby casting doubt on the extent to which de-identification is a credible method for using and deriving value from data while protecting privacy. Compounding the uncertainty is the fact that reidentification risks only increase as computing technologies become ever faster and data-centric products and services generate increasingly more data for linkage and analysis.

Thus, weak or unproven promises of de-identification are no longer acceptable to regulators around the world.

Proven techniques and processes like Anonos BigPrivacy are the minimum bar called for to support unlocking the value of data while respecting the rights of individuals. While no “silver bullet,” if implemented correctly dynamically applied de-identification can provide the technical operating instructions for both effective legal compliance and an operating system for respecting the information shared by individuals.

SUMMARY

Business
Business
Technology
Technology
Compliance
Compliance

The COVID-19 pandemic has established a "new normal" where the processing of digital assets to create timely data-driven insights is increasingly important.

The organisations that will re-emerge stronger are the ones that are driving digital transformation to maximise data value for the benefit of their customers.

Our premise is that organisations need **Speed To Insight, Lawfully and Ethically**. Without this speed to insight, organisations will be left behind, and without doing it lawfully and ethically, risk of liability or disruption to operations arises.

Anonos BigPrivacy software and API-based solutions uniquely embed policy, privacy and security controls into data flows to manage risk while data is in use.

Anonos solves one of the biggest challenges to successful digital transformation: balancing data utility and data protection to enable maximum value from sophisticated analytics, AI, machine learning (ML), data sharing, combining, and enriching in real-time, while ensuring that data is protected in use.

For an organisation to be successful, it must harmonise data risk management and use by fully involving and aligning different stakeholder groups within the organisation. No matter an individual's area of responsibility, they will need buy-in from other stakeholder groups to achieve demonstrable harmonisation of data protection and data innovation.

Accordingly, this Blueprint is organised into three parts to cover the perspectives of **BUSINESS, TECHNOLOGY** and **COMPLIANCE**. All three areas must work together and be aligned to achieve maximum data utility and protection simultaneously.

In the **BUSINESS** section we first discuss the key business benefits that arise from using Anonos BigPrivacy by focusing on "What We Solved For and Why" and discussing critical use cases that Anonos' software, API-based solutions and IP address. Then, we provide excerpts from a series of "Fireside Chats", with Doug Laney, the author of *Infonomics: How to Monetize, Manage and Measure Information as an Asset*.[1] These "Fireside Chats" approach the issue of organisational data use and risk management from several different perspectives and highlight a number of business use cases that Anonos helps to solve.

[1] See https://www.gartner.com/en/publications/infonomics

Next, in the **TECHNOLOGY** section we examine the technology behind Anonos BigPrivacy and how it functions, including a discussion of how GPDR-compliant Pseudonymisation is applied and how specialised privacy-respectful digital twins called “Variant Twins®” are created to manage risk.

Finally, the **COMPLIANCE** section covers a number of specific regulatory and legal benefits of using Anonos BigPrivacy technology, including what we call “Data Safe Havens”. These Data Safe Havens are specific GDPR legal benefits that come from dealing with data control management and data use enablement using BigPrivacy technology.

FOR A BRIEF OVERVIEW, WE SUGGEST REVIEWING THE FOLLOWING USE CASES:

CURRENT APPLICATIONS — Page

FUTURE APPLICATIONS — Page

BUSINESS

WHAT WE SOLVED FOR AND WHY

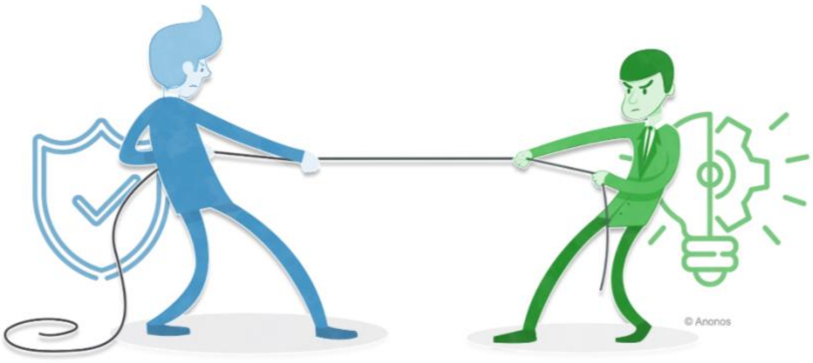
Anonos' technology and Intellectual Property (IP) future-proof **Speed To Insight, Lawfully & Ethically** *for decades to come* by enabling fully-protected batch and real-time use, sharing, combining and enriching of high risk and high-value multi-data asset ecosystems ("Big Data") on a global basis.

Embedding policy, privacy and security risk-based controls into data flows to protect both direct and indirect identifiers when data is in use enables sustainable speed to insight **while preserving 100% of source data value**. Anonos software and API-based solutions enable these capabilities. Alternatively, third parties may license Anonos IP to incorporate this functionality into their platforms to standardise technical interoperability processes for data movement and sharing.

From Inspiration to State-of-the-Art

Eight years ago, we saw a tug-of-war arising between (i) **data utility** and (ii) **data protection**. However, we also observed that traditional approaches to attempting to resolve this tug-of-war involved:

- Trying to obtain data subject consent;
- Anonymising high-risk data (which degraded accuracy and value);
- Restricting processing to controlled centralised environments; or
- Determining that desired uses were not lawful, with the subsequent deletion of high-value data.



While consent and anonymisation looked good on the surface, they had numerous limitations that businesses kept running into; centralised processing was not scalable; and deleting the data prevented valuable insights from being discovered.

So, we asked ourselves two key questions:

1. **Utility:** What if the value of data for secondary uses equalled or exceeded what was necessary for primary business operations?
2. **Protection:** What if laws or business best practices limited the unrestricted flow, collection and use of sensitive or regulated data?

Our Thesis

In 2012, we predicted[2] that increasing volume, velocity and variety of Big Data would require:

1. **Utility:** Dynamic in-use risk-based controls at a fine-grained level to ensure high utility decentralised processing.
2. **Protection:** New laws restricting consent-based use and requiring enhanced technical risk-based controls for lawful data repurposing, sharing, combining and enriching.

The Current Reality

Eight years later, many global organisations are "hitting a wall" and unable to achieve their digital insight goals because they are:

- Forced to delete data they want to process.
- Limited to lower value centralised processing applications.
- Unable to:
  - Process high value and high-risk data in decentralised applications;
  - Access valuable data they want to process;
  - Share data internally and externally as desired;
  - Combine data sources to maximise value; and
  - Enrich data as desired.

For organisations desiring to share, combine and enrich data to achieve *Speed To Insight, Lawfully and Ethically*, Anonos uniquely resolves the tug-of-war between data utility and protection while preserving 100% of data utility and enabling compliance with data protection laws.

[2] The co-founders of Anonos, 20-year successful business partners Gary LaFever & Ted Myerson, previously predicted the increasing risks from a shift in the financial services market to algorithmic financial securities trading. Their previous company, FTEN, invented new technology enabling real-time dynamic risk controls that fueled innovative lawful trading. NASDAQ acquired their former company for nine figures in 2010 and deployed its technology in over 100 financial markets around the globe. They founded Anonos in 2012 to bring their visionary insight to bear to address an even more significant market opportunity – lawful and ethical Big Data use. Having revolutionised the financial securities industry by inventing new technology to manage risks that arise only at the precise time that a securities trade occurs ("at-trade risk"), they set out to accomplish an even bigger goal: to invent new technology to control risks that only arise at the precise time that data is actually put to use in Big Data processing ("in-use risk"). Anonos invested eight years in Research & Development (similar to a biopharma company) to understand the architectural underpinnings and shortcomings of the then-current state-of-the-art technology, which only protected data: (i) when at rest (in storage) or when being transmitted (in transit), but did not protect data when actually in use as required for Big Data; or (ii) when in use for limited centralised environments, but did not protect data when in use for widespread decentralised environments as required for Big Data.

The following use cases showcase some of the benefits of using Anonos software, API-based solutions or Intellectual Property (IP).

USE CASE: Decentralised Data Analytics, AI, ML, Sharing, Combining & Enriching

As noted elsewhere in this Blueprint,[3] there are numerous situations in which consent suffers from serious limitations as a lawful basis for processing personal data under the GDPR. First, information provided to data subjects to describe processing must be specific and yet easy to understand. This creates serious issues when attempting to explain complex processes such as data analytics, AI, ML, sharing, combining, or enriching. These limitations are one of the reasons that Legitimate Interests exists as an alternate legal basis. However, lawful Legitimate Interests processing requires adequate technical and organisational safeguards to help protect the fundamental rights and interests of data subjects.

You may be familiar with Digital Rights Management ("DRM") techniques that are used by companies to limit the numbers of copies of data individuals can make, or how they can otherwise access, music, movies and other digital content. BigPrivacy employs DRM-like principles, but "stands DRM on its head" in a manner that we refer to as Privacy Rights Management® or PRM®.[4] Specifically, BigPrivacy's fine-grained risk-based controls enable the selective use and sharing of personal data with improved multi-stakeholder engagement, all without exposure to unnecessary privacy, security or degradation-of-value risks.

Traditional approaches to data protection use no longer effective "static" approaches to protection. As a result, supposedly protected data can be traced back to a data subject because persistent identifiers replace a given data element everywhere it appears. Searching for a persistent identifier that repeats within or across data sets can provide a malicious actor with enough information to unmask the identity of a data subject. Two well-known cases of such unauthorized reidentification involved AOL[5] and Netflix[6]. Over time, due to advances in technology and threat-actor sophistication, persistent identifiers can be ever more readily linked back to individuals via the "Mosaic Effect."[7] An oft cited example of the "Mosaic Effect" saw three seemingly "anonymous" data sets that used the same persistent identifiers – each composed of the zip code, age and gender of US citizens – combined to identify up to 87% of the population of the United States by name.[8]

In contrast, Anonos BigPrivacy uniquely combats the Mosaic Effect by using dynamic de-identifiers, as described in the TECHNOLOGY section below, to introduce uncertainty (entropy) at the data element level. From there, the data controller can selectively reveal only the level of identifiable data which a given user is authorized to see, at a specific time for a specific purpose. However, none of these

[3] See the discussion on "Limitations of Consent" in the TECHNOLOGY section, and the description of "Data Safe Have #2 - Legitimate Interests Processing" in the COMPLIANCE section, below.

[4] Ted Myerson, Co-Founder and President of Anonos, presented a **TED** Talk on how Privacy Rights Management (PRM) as enabled by Anonos BigPrivacy "stands DRM on its head." A video of, and the transcript for, this **TED** Talk is at https://www.ted.com/talks/ted_myerson_big_data_needs_big_privacy. "**TED** Talk" is a trademark of Ted Conferences, LLC.

[5] See https://techcrunch.com/2006/08/09/first-person-identified-from-aol-data-thelma-arnold/

[6] See https://bits.blogs.nytimes.com/2010/03/12/netflix-cancels-contest-plans-and-settles-suit/

[7] See https://www.MosaicEffect.com

[8] See http://dataprivacylab.org/projects/identifiability/paper1.pdf

advantages achieved through finer granularity prevent BigPrivacy from reproducing, when authorized, up to 100% of the value and utility of original source data.

Anonos' patented technology uniquely protects data dynamically so that decentralized analytics, AI, ML, data sharing, combining, and enriching can satisfy the requirements for Legitimate Interests processing under the GDPR. Anonos BigPrivacy and Lawful Insights API provide risk-based controls for data enablement, allowing organisations to step outside of the silos that centralised solutions create.

USE CASE: Compliant Cross Border Data Flows

The June 2020 World Economic Forum Whitepaper titled "*A Roadmap for Cross-Border Data Flows: Future-Proofing Readiness and Cooperation in the New Data Economy*"[9] ("WEF Whitepaper") highlights the following:

> *…Now more than ever, cross-border data flows are key predicates for countries and regions that wish to compete in the Fourth Industrial Revolution and thrive in the post COVID-19 era…*
>
> *…The security of data when it moves across borders is of fundamental concern to both companies and governments, both in terms of risk mitigation and security of proprietary data and intellectual property (IP). The absence of, or the risk of the absence of, security measures further undermines trust and produces friction for cross-border data sharing...*
>
> *…The use of open or standard application programming interfaces (APIs) for data sharing should be encouraged by governments to improve technical interoperability…*

As described in this Blueprint, Anonos BigPrivacy uniquely solves the biggest challenge to data sharing, combining and enriching for cross-border data flows: enabling maximum utility from data analytics, AI and ML in real-time, lawfully and ethically. This is because traditional centralised data protection technologies, including encryption, anonymisation, static tokenisation, and differential privacy:

- Significantly degrade the utility of data, distorting the accuracy and predictability of the derived insights;
- Fail to deliver effective protection against unauthorised reidentification in decentralised processing environments; and
- Limit the use of data for data sharing, combining and enriching.

The capability of Anonos BigPrivacy software, API-based solutions and IP to enable policy, privacy and security risk-based controls to be embedded **into** data flows enables organisations and countries to respond to the following calls-to-action in the WEF Whitepaper necessary for lawful and ethical cross-border data flows:

- Allowing data to flow by default;
- Establishing a new level of data protection;
- Prioritising Cybersecurity; and

[9] See https://www.weforum.org/whitepapers/a-roadmap-for-crossborder-data-flows-future-proofing-readiness-and-cooperation-in-the-new-data-economy

- Prioritising technical interoperability, data portability and data provenance.

USE CASE: IOT and 5G: Maximum Data For Digital Insights[10]

The Internet of Things (IoT) promises dramatic increases in the number and types of devices and sensors capturing data reflective of real-time situations, including valuable geolocation data and other information. 5G provides greater transmission speed, lower latency, and greater numbers of connected devices. The combination of IoT and 5G presents an unparalleled opportunity for valuable digital insight.

Unfortunately, due to ineffective in-use data risk management capabilities, privacy has historically been enforced using data collection minimisation – requiring that all the data that is not immediately needed must be deleted. For example, the US Federal Trade Commission (FTC) staff issued a 2015 report titled "*Internet of Things: Privacy and Security in a Connected World."*[11] This report recommended that the way to protect the privacy of IoT data was the wholesale deletion of information. This proposal was so absurd on its face that two FTC commissioners refused to endorse the report.[12] A non-partisan research firm (the Information Technology and Innovation Foundation or ITIF), highlighted problems with data collection minimization:

> *"The FTC's report correctly recognizes that the Internet of Things offers potentially revolutionary benefits for consumers and that the industry is still at an early stage, [but the report] unfortunately attempts to shoehorn old ideas on new technology by calling for broad-based privacy legislation"; further, "in calling for companies to reduce their use of data, the FTC misses the point that data is the driving force behind innovation in today's information economy."* [13]

So, how do we allow for the privacy-respectful collection and use of data made possible by IoT and 5G? A new approach to data protection is required: one that uses in-use risk management controls to enforce data **use** minimisation (versus data **collection** minimisation). Anonos BigPrivacy does exactly this by reducing risk associated with data in use (i.e., "de-risks" the data) anywhere in the data flow. BigPrivacy technologically enforces the use of only the minimum level of identifiable data necessary for each process to protect data on a dynamic per-use basis. This is the essence of data **use** minimization.

Data protection can be improved by allowing more data to be collected with appropriate technical and organisational safeguards. By collecting more data about individuals, sophisticated analytics, AI, and ML are possible using less-identifying data that might otherwise have been deleted (leaving only directly identifying data for desired processing). With effective in-use data risk management controls – like those provided by Anonos BigPrivacy – the increase in data collection enabled by IoT and 5G can actually improve privacy-respectful processing.

[10] Anonos is building out this capability by leveraging BigPrivacy technology and Intellectual Property (IP) rights with partners.

[11] See https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

[12] See https://itif.org/publications/2015/01/27/ftc%E2%80%99s-internet-things-report-misses-mark

[13] Id.

USE CASE: Blockchain: How to Make it GDPR Compliant[14]

The defining feature of blockchains is their integrity (i.e., the ability for users of network to trust the accuracy of the data stored in the blocks of the chain), which is guaranteed by their immutability. Once a block has been verified and added to the chain, it may not be removed, edited, or updated. Modifying the data stored in any one block would 'break' (i.e., invalidate) all the downstream blocks in the chain. While blockchain data, in the vast majority of cases, is protected by encryption or static tokenisation, it is easy to envision cases where individuals will want to exercise their "right to erasure/right to be forgotten" pursuant to the GDPR by requesting that their data be removed from the blockchain. With public blockchain platforms, such a request would not be possible to fulfill without destroying the integrity of the entire chain.

The Financial Conduct Authority (FCA), the financial regulatory body for the United Kingdom, has warned firms developing blockchain technology to beware of the incompatibility between immutability and the GDPR.[15] Some solutions to this issue have been proposed, such as allowing administrators to edit a blockchain where necessary. However, allowing editing of a blockchain destroys the defining characteristic of blockchain: integrity guarantees delivered via immutability.

The GDPR was authored under an assumption that custodians of data would continue to be centralized entities and thus doesn't account for decentralized systems such as blockchain.

BigPrivacy uniquely enables blockchain and other distributed ledger technologies (DLTs) to comply with GDPR requirements like the "right to erasure/right to be forgotten" while still satisfying the immutability, auditability, and verification requirements mandated by DLTs for decentralized storage of transactional data.[16]

See Appendix A for more information on the Anonos Patent Strategy and Portfolio.

[14] Anonos is building out this capability by leveraging BigPrivacy technology and Intellectual Property (IP) rights with partners.

[15] See https://www.fca.org.uk/publication/feedback/fs17-04.pdf

[16] See Anonos granted Patent No. 10,572,684 titled "*Systems and Methods for Enforcing Centralized Privacy Controls in De-Centralized Systems*." See Appendix A for more information on the Anonos Patent Strategy and Portfolio.

ONLINE FIRESIDE CHATS

Broader business benefits from **Speed To Insight, Lawfully & Ethically** are highlighted in the following excerpts from a series of “Online Fireside Chats” between Doug Laney, Data & Analytics Strategy Principal at Caserta[17] and author of the book “*Infonomics: How to Monetize, Manage and Measure Information as an Asset*,” and Gary LaFever, CEO and General Counsel at Anonos, available at www.SpeedToInsight.com.

- First, these “Online Fireside Chats” look at what **Speed To Insight** is and why it matters, and why a shift away from traditional centralised data protection technologies is critical to achieving this goal.
- Next, they cover some of the challenges involved in maximising data value, such as regulatory controls surrounding both PII and non-PII personal data, over-reliance on traditional data protection technologies, and a lack of access to data.
- Then, they examine how organisations can move toward gaining frictionless insights from their data with a more automated and risk-based control approach to data use and protection, including in decentralised processing environments. This includes data sharing, combining and enriching, as well as machine learning and AI data use.
- Finally, they discuss the shift in expectations for data protection that has come with the GDPR, and what this means for old data uses that may no longer be lawful. How can organisations defend their previous uses of data, or maintain the ability to use pre-GDPR data that is still valuable?
- These discussions are supported by infographics and diagrams that display some aspects of how BigPrivacy functions, which will be expanded upon in the TECHNOLOGY section below.

Speed To Insight in the New Normal

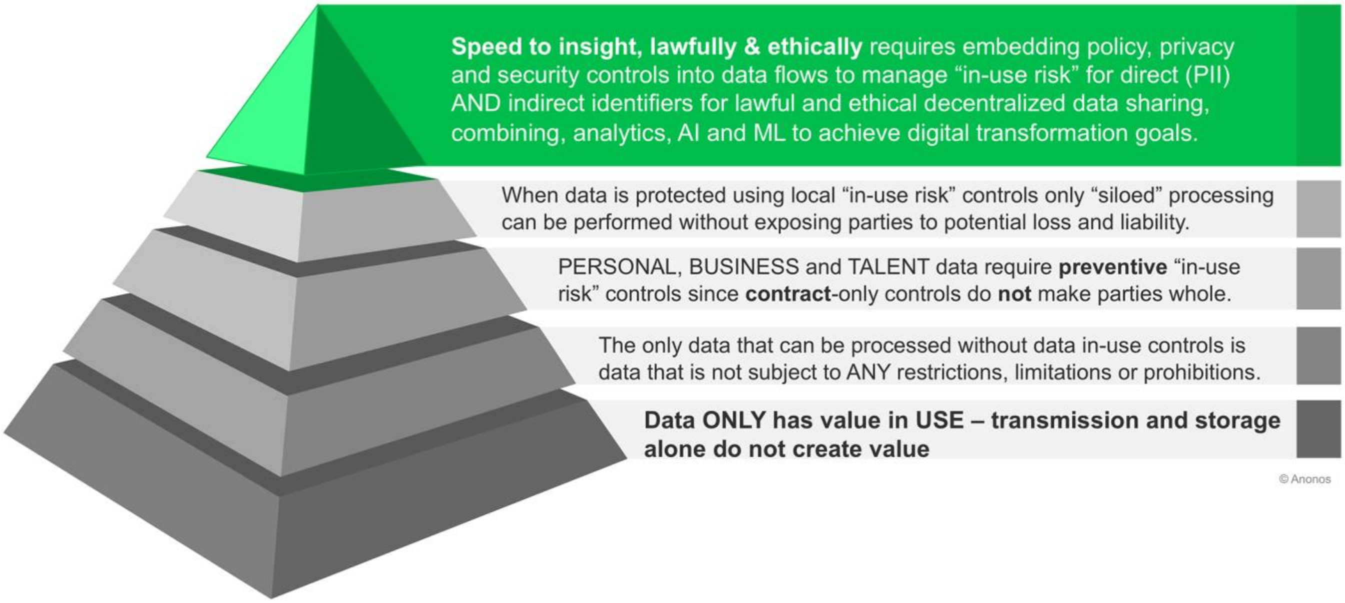
DOUG LANEY: Gary, I think the COVID-19 pandemic is producing a “new normal” where the processing of digital assets to create timely data-driven insights is increasingly important. One need only look at the impact of consumers not being able to visit brick-and-mortar stores for months, resulting in an extraordinary increase in the use of digital payments. This is proof of an increasingly savvy digital customer base. Organisations that effectively leverage digital insights to provide customers with context-aware, personalised offerings will be the winners in this new normal.

This leads me to believe there will be little middle ground between data insight “haves” and “have-nots.” Organisations that cannot implement sustainable strategies for developing and refining digital insights run the risk of becoming non-competitive. In contrast, organisations that implement sustainable, trustworthy and transparent data insight strategies will thrive. Successful data use, sharing and combination arrangements between partners will be the difference between winners and losers.

GARY LAFEVER: I completely agree. The overwhelming increase in people working from home and purchasing goods online has dramatically accelerated our transition to a largely digital world. To survive and thrive, organisations need data-driven insights to anticipate and react to quickly changing buying patterns. This shift underscores the importance of moving beyond traditional approaches to

[17] Caserta (https://caserta.com/) is a strategic consulting and innovation technology implementation firm that helps clients leverage emerging technologies to advance business leadership.

data protection to support new requirements for businesses to gain “Speed To Insight”, but critically, organisations need those insights to be "Lawful and Ethical” as well.



Speed to insight, lawfully & ethically requires embedding policy, privacy and security controls into data flows to manage “in-use risk” for direct (PII) AND indirect identifiers for lawful and ethical decentralized data sharing, combining, analytics, AI and ML to achieve digital transformation goals.

When data is protected using local “in-use risk” controls only “siloed” processing can be performed without exposing parties to potential loss and liability.

PERSONAL, BUSINESS and TALENT data require **preventive** “in-use risk” controls since **contract**-only controls do **not** make parties whole.

The only data that can be processed without data in-use controls is data that is not subject to ANY restrictions, limitations or prohibitions.

Data ONLY has value in USE – transmission and storage alone do not create value

© Anonos

Data only has value when it is in use. Security technologies—like encryption—remain important for protecting data at rest and in transit, but they do nothing to generate digital value or create insights. When data is put to use, the protections afforded by security technologies no longer apply, because these technologies protect data only when in transit or at rest.

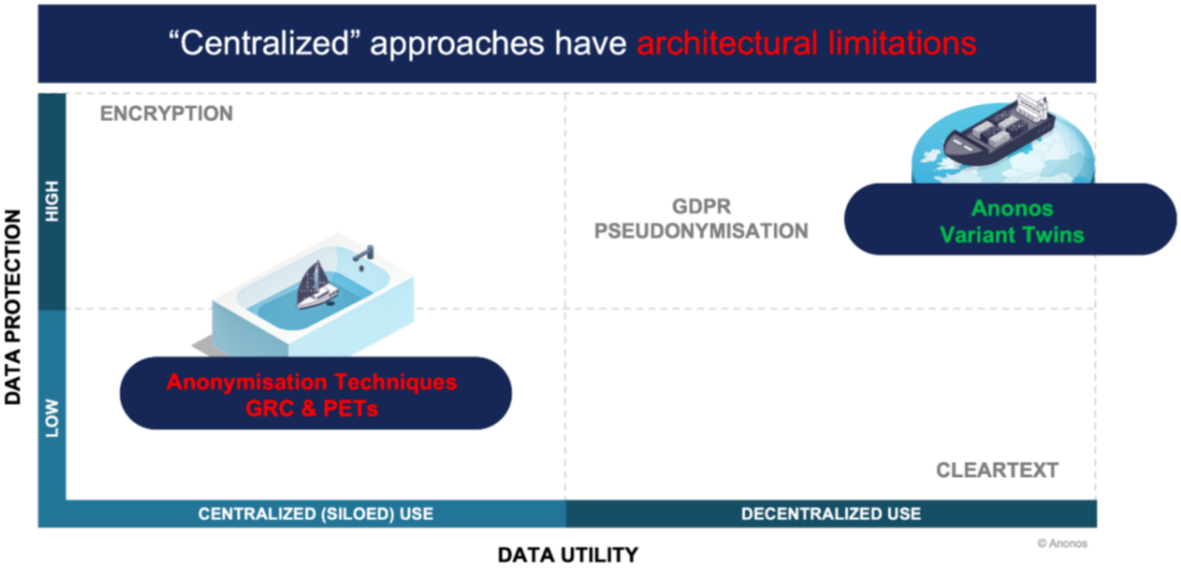
Traditional approaches to data protection also create tensions between the business desire to generate digital insights, and the obligation of security and privacy teams to protect their organisation against threats, liability and disruptions to operations from data misuse. While an organisation may be able to spin up a new cloud server in a few minutes, they may have to wait weeks or months to get security and privacy sign-off before going live with a new application on the server. The only data that can be safely used without security and privacy sign-off is data that is not subject to any restrictions.

Development teams focused on using Analytics, AI, ML and data-sharing, combining and enriching technologies to deliver desired business results without addressing data security and privacy risks expose their organisations to significant liability and potential disruption to operations.

This failure to comply with laws, rules and regulations applicable to high-risk but high-value data like personal, business and talent data is a misalignment that can lead to missed business opportunities.

Traditional data protection technologies, like anonymisation via tokenisation, generalisation or suppression, as well as newer techniques like Differential Privacy, synthetic data and homomorphic encryption, protect data when in use but *only* for centralised processing.

Choose Data Protection for Your Data Strategy



These traditional privacy protection techniques do not support decentralised data processing, sharing, combining, or enriching. Examples of desired decentralised processing include when you want to share or combine datasets between organisations, combine multiple datasets of your own, or use datasets from different places for machine learning and AI. Since traditional data protection technologies are centralised, they limit the availability of data needed to generate robust digital insights.

In addition, traditional centralised data protection technologies often only focus on protecting immediately identifying data, often referred to as Personally Identifying Information, or PII. But, recent laws like the California Consumer Privacy Act (CCPA) and the EU General Data Protection Regulation (GDPR) require protection of more than just PII. These laws extend the obligation for data protection to indirectly identifying data, such as age, gender, birthdate and location. When these indirect identifiers are combined, they can be used to re-identify an individual.[18] This is why laws like the CCPA and GDPR require their protection as well.

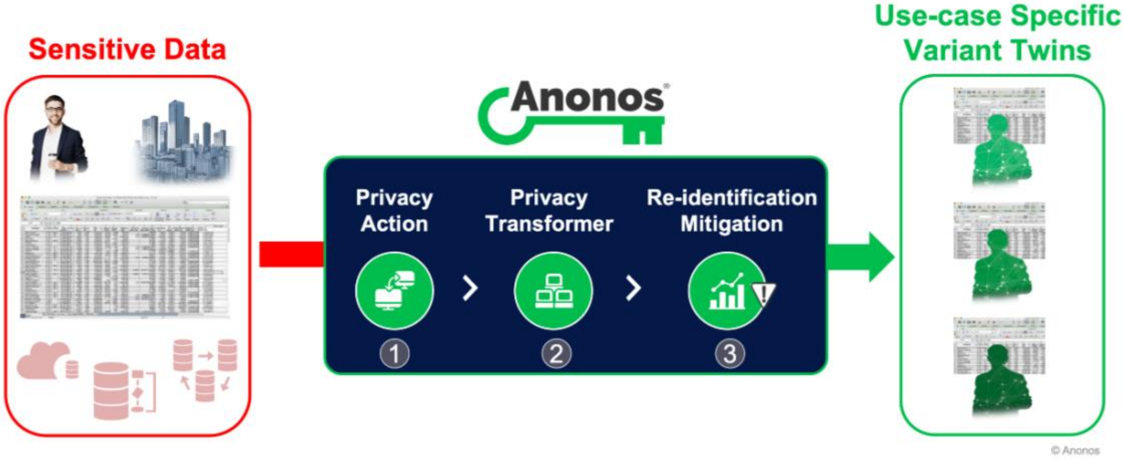
In summary, traditional centralised data protection technologies can:

1. Create insurmountable tensions between business and security/privacy teams;
2. Delay access to desired processing until digital insights are less timely and less relevant; and
3. Limit data insights to those available from centralised applications that cannot be linked together.

In contrast, Anonos decentralised data protection helps to resolve these issues by creating pre-approved schemas for non-identifying versions of data, called Variant Twins.

[18] See USE CASE: Decentralised Data Analytics, AI, ML, Sharing, Combining & Enriching above.

Anonos Variant Twins – Enable Lawful Data Use



Variant Twins can be created for different processes to selectively disclose only the level and type of data approved in advance by security and privacy teams for each use case. By embedding policy, privacy and security controls into data flows to manage risk**,** use-case specific Variant Twins enable lawful and ethical decentralised data use, sharing, and combining so that businesses can gain “Speed To Insight, Lawfully & Ethically.”

DOUG LANEY: Gary, can you provide a use case where Anonos technology helps to enable speed to insight, lawfully & ethically?

USE CASE: Speed To Insight, Lawfully & Ethically

GARY LAFEVER: Let's take the example of a global firm with EU employees that wants to do Talent Analytics around the world. Firms are only just starting to become aware that talent data must now be processed differently to remain lawful and to avoid undesirable disruptions to business operations.

These challenges arise primarily because:

- PII (as well as non-PII data that can become identifying when combined) creates liability if processed by employers based on the consent of EU employees because of the imbalance of negotiating power between the parties. This imbalance removes consent as an available basis for lawful processing of data for Talent Analytics under the GDPR.
- Similar problems can arise when sophisticated analytics, AI or ML are desired using non-employee Personal Data beyond the scope of what was described in detail to data subjects at the time of initial data collection.
- In addition, both PII and non-PII can cause significant disruption to operations when data subjects demand that all of their data (not just PII) be deleted or alternatively no longer shared with third parties. Data assets cannot be processed effectively when their **very** composition and availability change from day to day.

Anonos technology is different from other solutions. Centralised privacy enhancing technologies do not embed risk-based controls that flow **with** the data outside the centralized environment and so may not provide adequate protection to satisfy the balancing of interest requirements necessary for sophisticated analytics, AI and ML to be lawful.

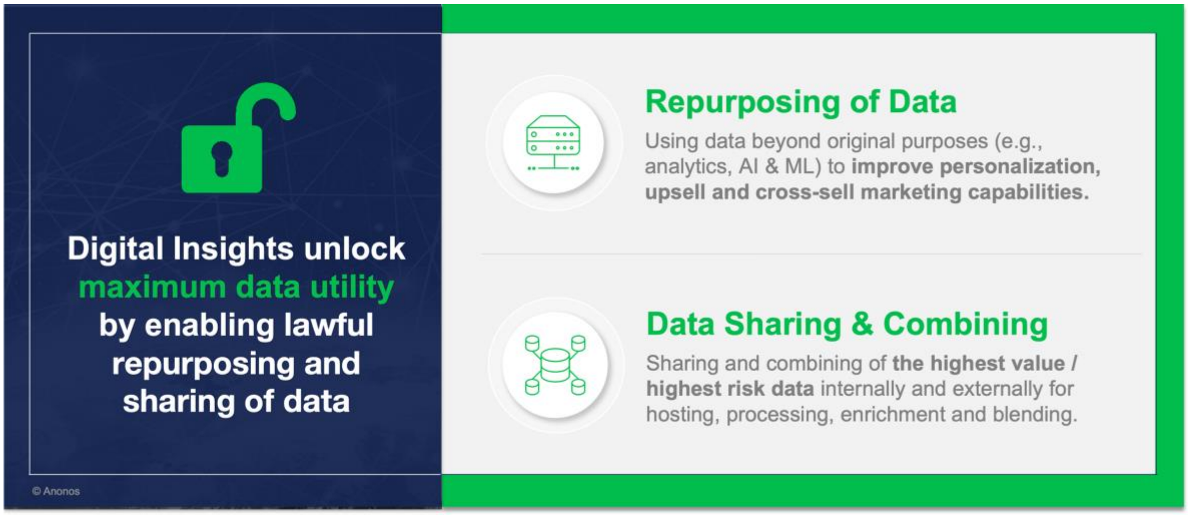
In contrast, Anonos decentralised data protection technology manages risk differently based on the level and nature of risk involved in different processes, regardless of where the data goes. This helps to ensure that digital insights are lawful and equitable, both within and between organisations.

Maximising Data Utility

DOUG LANEY: Infonomics, the term I coined in my book to describe monetising, managing, and measuring information for competitive advantage, is even more important for organisations to survive and thrive in today’s economy. They must maximise the value and pace of their data journey to succeed. Infonomics in these highly uncertain times does not tolerate inefficiency.

Successful data sharing and data use arrangements will make all the difference between post-pandemic winners and losers. Sustainable, trustworthy and transparent data strategies require that adequate risk-based controls exist to satisfy both customers and data sharing partners that the data collaboration will be safe as well as effective. Without adequate risk-based controls, organisations may not get access to the data they need for successful digital insights.

Top Global Use Cases



Digital Insights unlock maximum data utility by enabling lawful repurposing and sharing of data

Repurposing of Data

Using data beyond original purposes (e.g., analytics, AI & ML) to **improve personalization, upsell and cross-sell marketing capabilities.**

Data Sharing & Combining

Sharing and combining of **the highest value / highest risk data** internally and externally for hosting, processing, enrichment and blending.

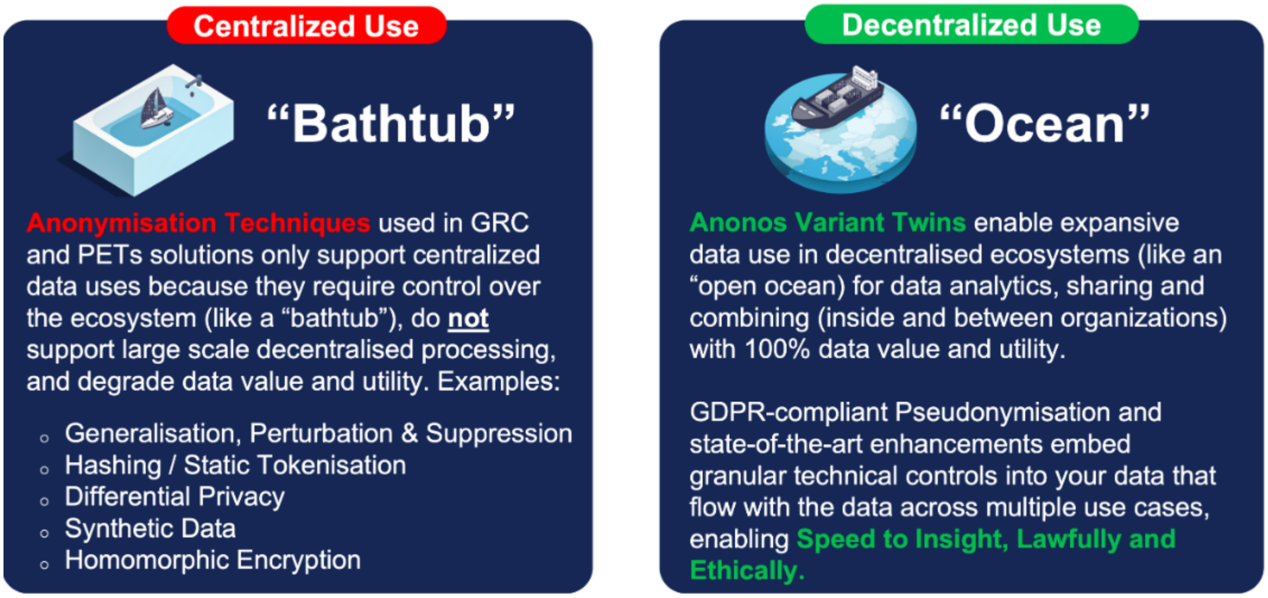
© Anonos

GARY LAFEVER: I think that’s right: effective extraction of digital insights requires risk-based controls that protect data in use while preserving utility.

Organisations can capture the most value when they can use data across multiple environments with different partners to discern insights. Maximum data value often comes from combining data sets, adding new information, and processing data from different perspectives. For effective digital insight strategies, organisations must control the exchange of specific information with different partners, meet contractual obligations, and process data under controlled conditions internally and externally.

Customers as well as data sharing partners may be unwilling to share data if the organisation who wants to collaborate with them does not have adequate risk-based controls in place.

What is Your Data Protection Doing for You?



Centralized Use

"Bathtub"

Anonymisation Techniques used in GRC and PETs solutions only support centralized data uses because they require control over the ecosystem (like a "bathtub"), do **not** support large scale decentralised processing, and degrade data value and utility. Examples:

- Generalisation, Perturbation & Suppression
- Hashing / Static Tokenisation
- Differential Privacy
- Synthetic Data
- Homomorphic Encryption

Decentralized Use

"Ocean"

Anonos Variant Twins enable expansive data use in decentralised ecosystems (like an "open ocean) for data analytics, sharing and combining (inside and between organizations) with 100% data value and utility.

GDPR-compliant Pseudonymisation and state-of-the-art enhancements embed granular technical controls into your data that flow with the data across multiple use cases, enabling **Speed to Insight, Lawfully and Ethically.**

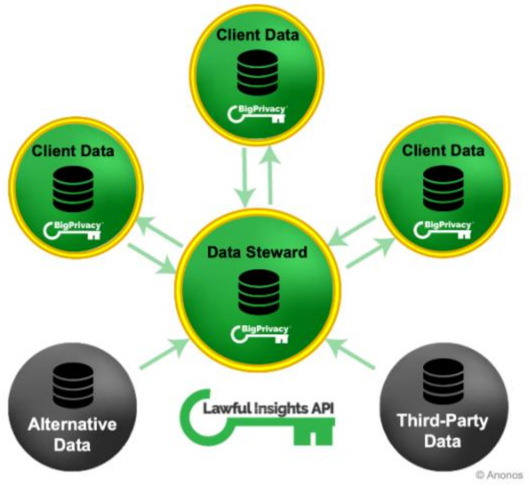
On the flipside, if an organisation relies on traditional data protection technologies, like anonymisation via tokenisation, generalisation or suppression as well as newer techniques like Differential Privacy, synthetic data or homomorphic encryption, they may not get access to all the data they need for effective digital insights. This is because these technologies only protect data for limited centralised purposes.

Anonos decentralised data protection technology resolves these issues by controlling data use risk across multiple environments and with different partners. This is helpful when organisations experience trouble getting access to third-party data to augment the value of their data assets, or when third parties express concern about potential liability or business disruption risk from using data provided by the organisation.

Access to Data for Digital Insights

Lawful Insights™ support disparate data flows based on capabilities of data sources (Client Data, Alternative Data, Third Party Data) to satisfy **Data Protection by Design and by Default** (GDPR) and **heightened de-identification** requirements (CCPA):

- **Inbound**: Lawful Insights API enables data sources to send encrypted data to a Data Steward that is immediately transformed into a **privacy respectful Variant Twin.**
- **Outbound**: Lawful Insights API enable a Data Steward to retrieve **privacy-respectful Variant Twins** that have been prepared by data sources.



BigPrivacy® maximizes data utility and value while preserving 100% accuracy by embedding policy, privacy and security controls into data flows to manage "**in-use risk**" for lawful and ethical decentralized data sharing, combining, analytics, AI and ML through leveraging:

- **Record-level dynamic pseudonymisation** that supports **relinking to all source data,** not just reversal of pseudonyms.
- **Microsegment (mSeg) pseudonyms** that support privacy respectful data enrichment and **omnichannel personalization without** surveillance.
- **Dynamic de-identifiers** within and between datasets to **defeat linkage attacks**.

Anonos decentralised data protection technology enables lawful and ethical multi-party processing both inside and outside of an organisation's environment. The key is our combination of anonymisation capabilities, GDPR-compliant Pseudonymisation (CCPA-compliant heightened de-identification), and patented risk-based controls that flow with the data to control relinking capabilities.

By embedding dynamic, risk-based controls into data flows, as highlighted by Gartner when they awarded Anonos "cool vendor" status, Anonos technology is able to maximise data utility and value while preserving 100% accuracy.[19]

DOUG LANEY: Gary, can you provide a use case where Anonos technology helps to enable access to data for digital insights?

USE CASE: Enabling Access to Data for Digital Insights

GARY LAFEVER: Let's look at a consulting firm that wants to provide supporting data and analysis to help clients make and execute strategic and tactical decisions. Professional advisory firms are increasingly concerned about lawful and ethical access to the data that is necessary for them to provide value to clients. Challenges arise primarily because:

- The data obtained from customers may be problematic in terms of liability.
- Dealing with PII, or non-PII data that can **become** identifying when combined with other data, can cause disruptions to operations. This occurs when data subjects demand that their data be deleted, as firms must then show that it is deleted. Data assets cannot be processed effectively when they are constantly changing.
  - Value-add to data is needed, e.g. combining data among clients, and from third parties.
  - Third-party data can often be "contaminated" with PII.

[19] See https://www.anonos.com/awards

- – Alternative data sources may not be willing to accept the liability and risk associated with sharing data.

Anonos decentralised data protection is different from other solutions. Tokenisation and other PETs work in some circumstances, but do not embed risk-based controls that flow with the data outside centralized environments to enable decentralised processing. Anonos technology does not rely on privacy "boxes" or "cages" to protect data: rather, we put the risk-based controls into the data, so that risk is managed wherever the data goes, even during data sharing, combining, or transforming.

This approach ensures uninterrupted access to third-party data and alternative data and protects consulting firms from legal or operational risk. Consulting firms can:

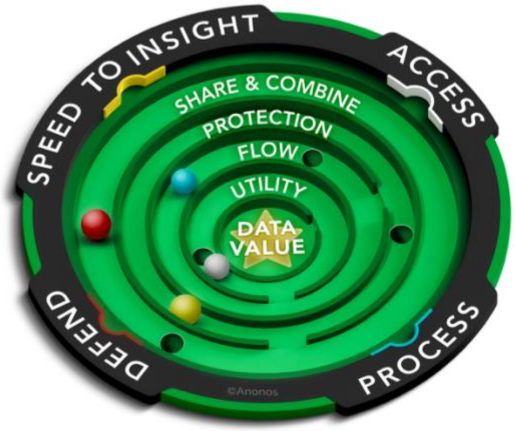
- Use Lawful Insights API™ to transform data coming from third-party data providers into usable "Variant Twins" that are lawful, ethical, and accurate.
- Enable alternative data sources to use Lawful Insights API to control the data they make available to the consulting firm or its clients.
- Provide customers with access to BigPrivacy to send data to and receive data from the consulting firm for specific purposes.
- Share and combine data with and between clients with reduced risk of liability or operational disruption for any party.

Reducing Friction in Data Processing

DOUG LANEY: Gary, what does Anonos do to reduce friction in data processing, while complying with the requirements of both internal and external parties? Can data be made available to achieve business outcomes across different ecosystems in a way that is lawful and ethical, but also efficient and timely?

GARY LAFEVER: Anonos' decentralised data protection technology allows data to be processed in an automated way to achieve desired business outcomes with full awareness of, and the ability to remove, potential roadblocks to processing. Anonos does so by embedding policy, privacy and security risk-based controls into data flows. These are centralised controls over decentralised processing, that automate the balancing of complex, multi-issue processing to comply with established policy, privacy and security requirements.

Anonos Simplifies Complex Processing to Maximize Data Value



Anonos automates the balancing of complex, multi-issue processing in compliance with established policy, privacy and security controls to manage "**in-use risk**" for lawful and ethical decentralized data sharing, combining, and analytics.

Data sharing, combining and enriching is where data value, insight, privacy and security meet. Without Anonos decentralised data protection, desired data uses may be too risky or unlawful, or the value of the data may be diminished. Processing performed using traditional centralised data protection technologies may be too slow and inefficient to obtain digital insights as quickly as they are needed. With Anonos, however, organisations can comply with internal and external requirements while maximising data control, use, and value.

Once an organisation builds a portfolio of desired processes using Anonos technology configurations, Data Protection Impact Assessments (DPIAs) and related technical and organisational safeguards, new use cases can be supported using variations of these, so that bespoke DPIAs become the exception.

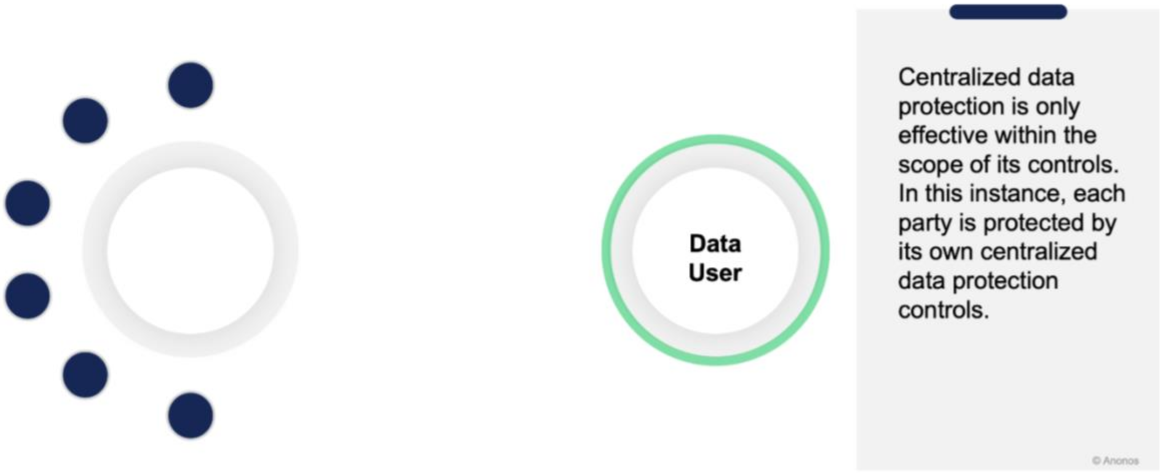
Anonos does this by leveraging:

- Record-level GDPR pseudonymisation and CCPA-heightened de-identification to support controlled relinking to all source data, not just reversal of pseudonyms or de-identifiers;
- Microsegment (mSeg™) level pseudonyms/de-identifiers to support privacy-respectful data enrichment and omnichannel personalisation that does not require surveillance of individuals; and
- Dynamic de-identifiers within and between datasets to defeat linkage attacks.

DOUG LANEY: Gary, can you provide a use case where Anonos technology helps to enable automated processing for frictionless insights?

USE CASE: Automated Processing for Frictionless Insights

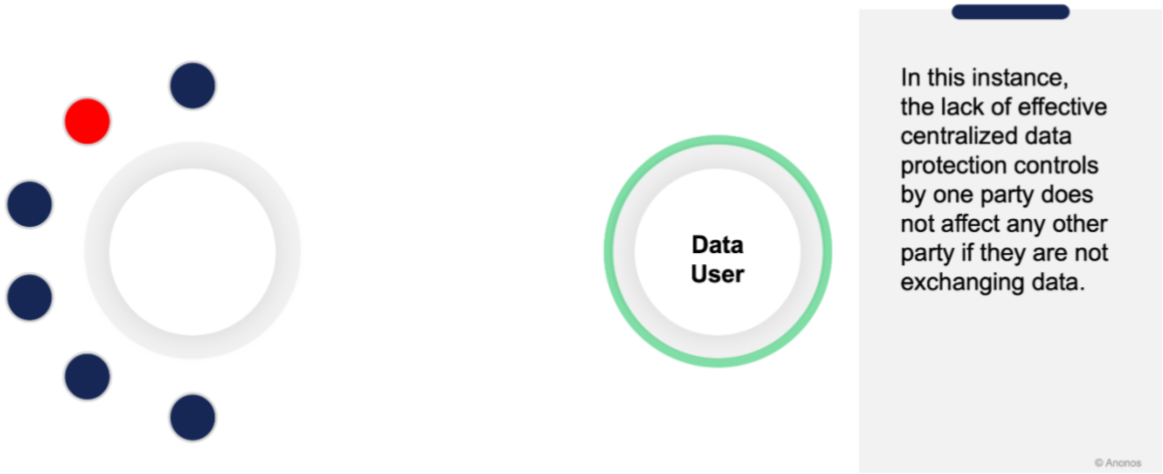
GARY LAFEVER: Nothing creates more friction between a Data User and a Third-Party Data Provider than liability from tainted or unlawful data or disruptions to operations when a data subject requests that their data be deleted. The following graphics highlight differences between traditional centralised approaches to protecting data and Anonos' decentralised approach. Anonos enables organisations to collect, use and share data in more efficient, faster, and more focused ways, with an entirely new approach.



Centralized data protection is only effective within the scope of its controls. In this instance, each party is protected by its own centralized data protection controls.

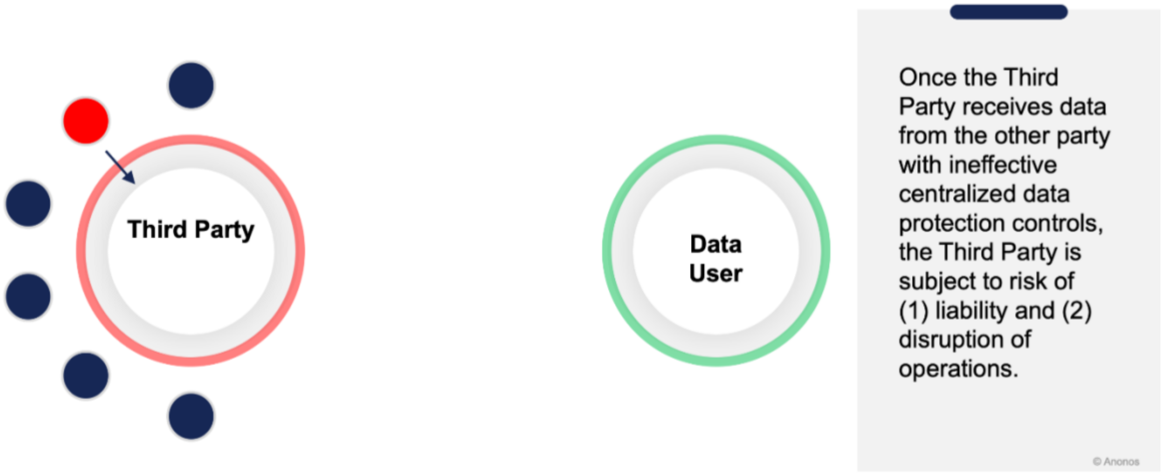
The dark blue circles on the left represent parties holding data that they have protected using their own centralised data protection like anonymisation via tokenisation, generalisation or suppression, or newer techniques like Differential Privacy, synthetic data or homomorphic encryption, all of which protect data for centralised processing only.

Here, the Data User is protected from risk.



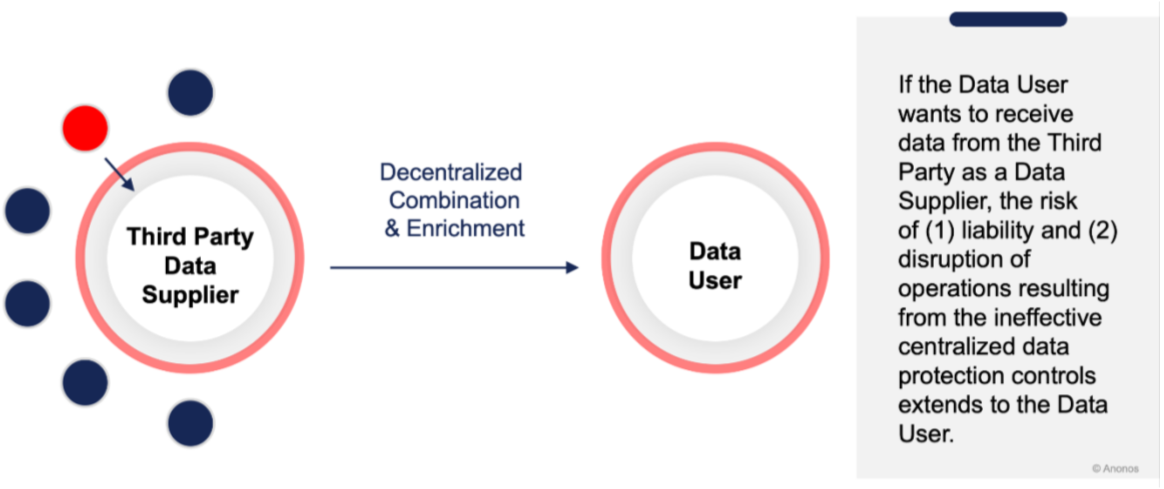
In this instance, the lack of effective centralized data protection controls by one party does not affect any other party if they are not exchanging data.

In this example, none of these parties are sharing data with each other. So, even if one party has defective controls, it does not affect other parties since there are no links between them.



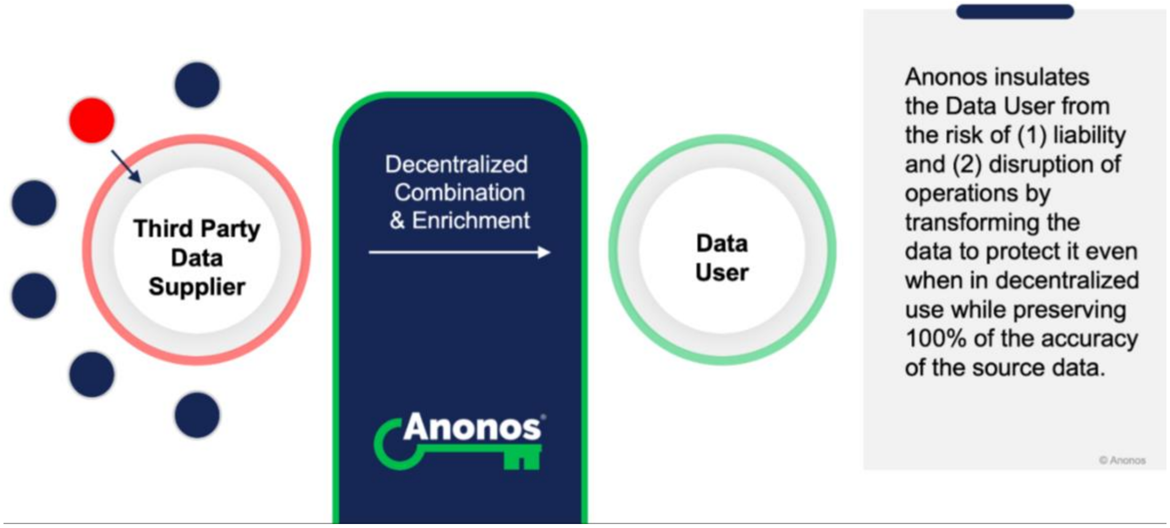
Once the Third Party receives data from the other party with ineffective centralized data protection controls, the Third Party is subject to risk of (1) liability and (2) disruption of operations.

But if a Third Party receives data from the external party with defective controls, that Third Party is now exposed to potential liability and disruption of operations.


Third Party Data Supplier
Decentralized Combination & Enrichment
Data User
© Anonos

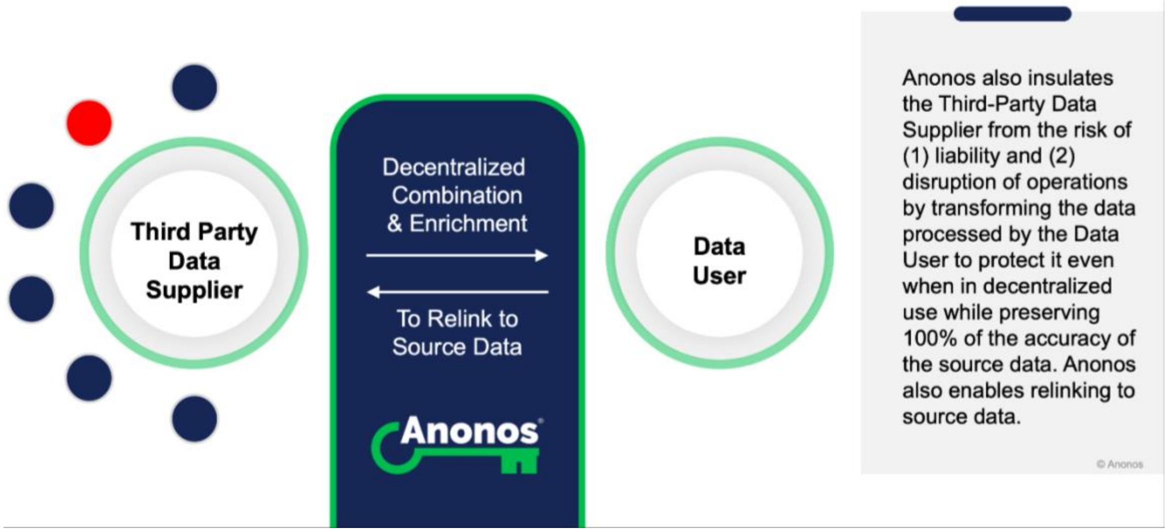
If the Data User wants to receive data from the Third Party as a Data Supplier, the risk of (1) liability and (2) disruption of operations resulting from the ineffective centralized data protection controls extends to the Data User.

If that Third Party then shares data with the Data User, the Data User is now also exposed to potential liability and disruption.


Third Party Data Supplier
Decentralized Combination & Enrichment
Anonos
Data User
© Anonos

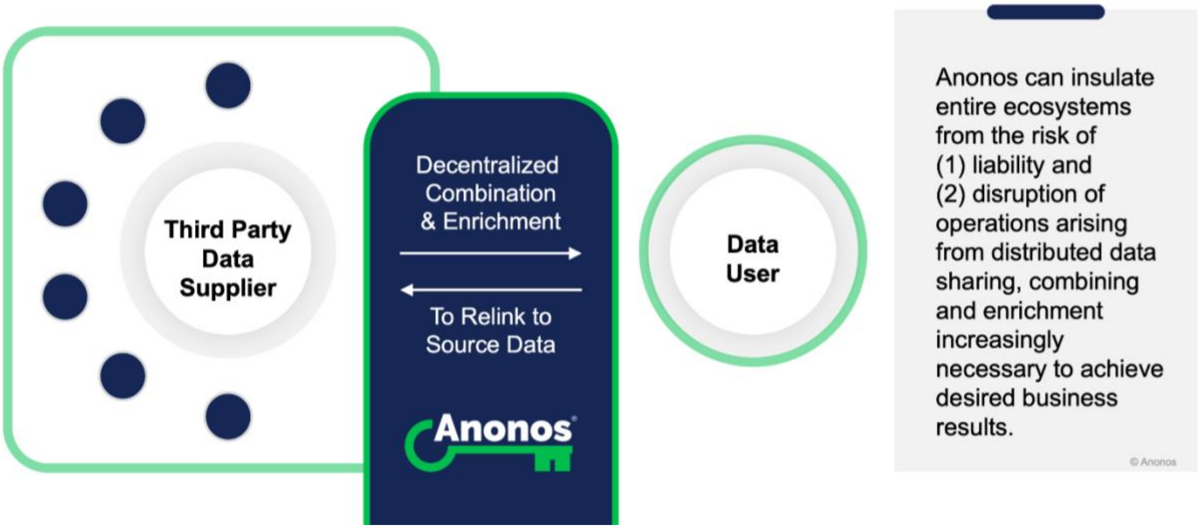
Anonos insulates the Data User from the risk of (1) liability and (2) disruption of operations by transforming the data to protect it even when in decentralized use while preserving 100% of the accuracy of the source data.

Anonos decentralised data protection technology is specifically designed for this situation and insulates the Data User from these risks.



Anonos also insulates the Third-Party Data Supplier from the risk of (1) liability and (2) disruption of operations by transforming the data processed by the Data User to protect it even when in decentralized use while preserving 100% of the accuracy of the source data. Anonos also enables relinking to source data.

If the Third Party wants to use the results of processing by the Data User to relink to source data, Anonos technology also insulates the Third Party from potential liability and disruption of operations in the other direction.



Anonos can insulate entire ecosystems from the risk of (1) liability and (2) disruption of operations arising from distributed data sharing, combining and enrichment increasingly necessary to achieve desired business results.

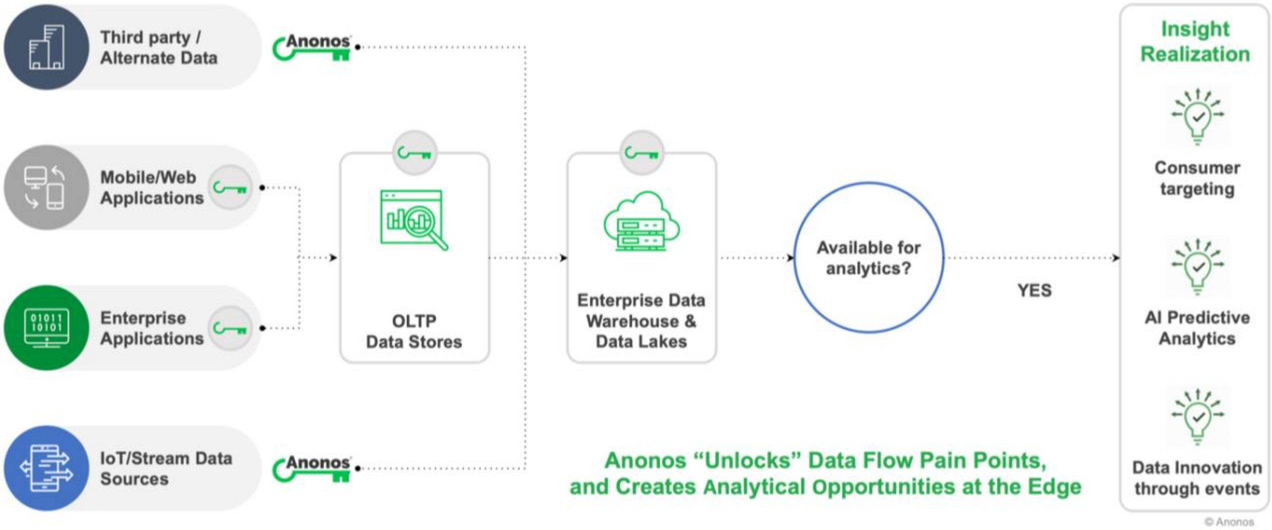
Anonos patented decentralised data protection technology enables businesses to achieve desired outcomes with confidence, by protecting entire data ecosystems in both directions.

Defending Digital Insights

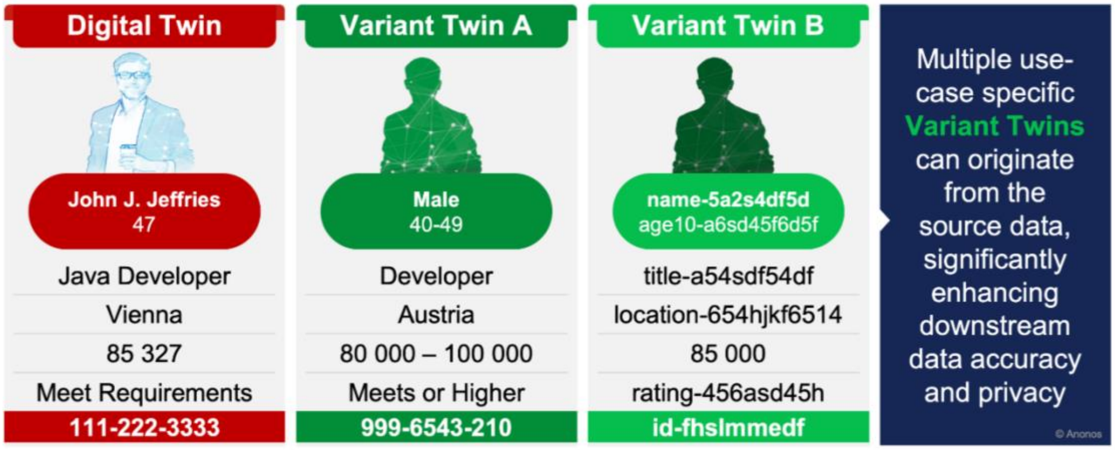
DOUG LANEY: Some people wonder why digital insights are so important in the post-pandemic world. It's because organisations need new sources of value to replace those that have been lost, and they also need to diversify revenue streams. In addition, they need to solidify relationships with existing customers and improve upon partnerships via commercial data exchange relationships.

It seems to me that Anonos can help to reconcile the opposing forces of regulation and monetisation necessary to achieve these results. On the one hand organisations have new controls and regulations to contend with, which reduces their ability to derive digital insights from leveraging data assets both internally and externally. On the other hand, they have an imperative to monetise data in new and innovative ways to generate new value streams. Gary, can you explain how Anonos helps to bridge this widening divide?

GARY LAFEVER: Anonos data protection technology can be integrated into data lakes, into streaming data, into batch processing, or applied at the network edge, to enable decentralised data use, sharing, combining and enriching to be accomplished in a flexible and scalable way.



This means that digital insights can be achieved in compliance with global data protection standards, vertical industry regulations and data sovereignty or localisation requirements. Anonos does this by combining the benefits of anonymisation, Pseudonymisation (as newly defined in the GDPR), CCPA-heightened de-identification and patented proprietary techniques. Using this combined approach, Anonos enables accurate decentralised data use, sharing, combining and enriching in compliance with regulatory and risk management requirements.



The core of Anonos' capabilities is centered around our non-identifying versions of data, called Variant Twins, which are our patented approach to controlling selective date use and disclosure risk.

With Variant Twins, organisations only provide the type and level of identifiable data needed for each authorised process. This aligns with modern data protection principles like data minimisation and purpose limitation. Because all Variant Twins are derived from the original source data, organisations suffer no degradation in data value or accuracy.

Variant Twins:

- Deliver resistance to reidentification of anonymous data;
- Enable data controllers to retain absolute control over the re-linkability of their data;
- Preserve 100% of the utility of source data;
- Protect data in use; and
- Activate express statutory benefits under data protection laws.

Variant Twins improve upon the statutory capabilities and benefits allocated to state-of-the-art privacy techniques like Pseudonymisation under the GDPR and heightened de-identification under the CCPA. They do so in a way that enables digital insight objectives to be achieved while respecting and enforcing the fundamental rights of data subjects in a data-driven ecosystem.

DOUG LANEY: Gary, can you share a use case where Anonos technology helped to defend the lawfulness and ethics of desired processing”?

USE CASE: Defending Lawfulness and Ethics of Desired Processing

GARY LAFEVER: We are working with a European bank that is interested in using enhanced market intelligence to develop new cross-sell and upsell opportunities for existing and prospective customers. But to create useful market intelligence to generate new opportunities for clients and partners, the bank required access to timely digital insights and expanded data use.

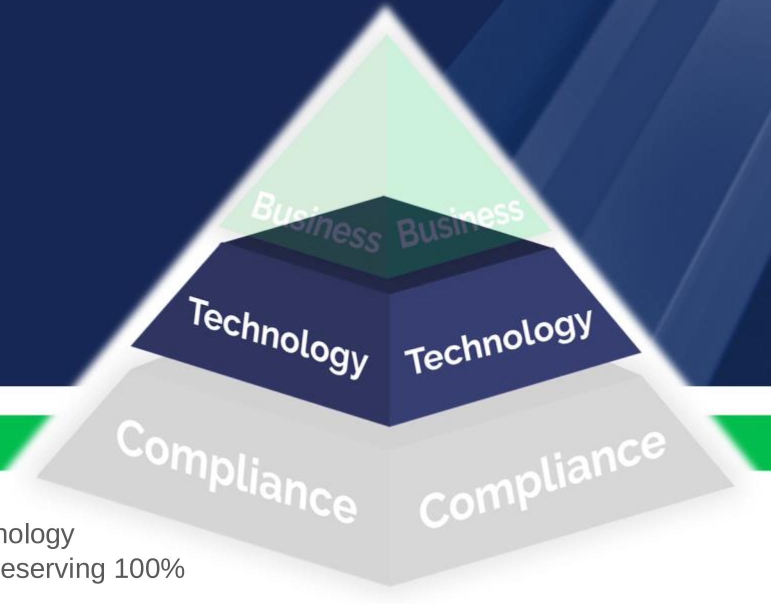
Some of the challenges were:

- Sharing and exchanging data with partners and third parties to better serve customers and partners was necessary to create better business outcomes.
- Customer needs and expectations could not be met or understood without the data.
- Access to the data, processing and the resulting insights were not lawful without enhanced data protection capabilities being put in place, but the protection could not reduce data utility or enablement.

By leveraging BigPrivacy Variant Twins, the bank is able to preserve data value while enabling sophisticated risk-based decentralised processing so that it can benefit from:

- Access and use of data that they otherwise would not have.
- Enhanced accuracy in insights and market intelligence.
- Data exchange with partners for more effective offerings to customers.
- Increased availability, and improved stability, of data assets.

TECHNOLOGY



Anonos decentralised data protection technology enables lawful repurposing of data while preserving 100% of the source data value.

Anonos allows organisations to maximise data utility and expand their opportunities to ethically process, share, combine and enrich data in compliance with data privacy and data protection regulations.

This section will first cover the foundations of Anonos BigPrivacy from a logical and technical perspective, an approach called functional separation.

We will then address some of the shortcomings of the lawful basis of consent when used by itself under the GDPR, and why technological solutions are needed to handle non-consent bases for processing data lawfully.

Next, we look at GDPR-defined Pseudonymisation as a new approach to dealing with the protection of personal data. We examine the implementation of Pseudonymisation as an outcome in the BigPrivacy solution and compare it to traditional anonymisation approaches.

Finally, we explain how Variant Twins leverage Anonos patented Controlled Linkable Data[20] and k-anonymity risk testing and conclude with a practical guide to using Anonos BigPrivacy to create Variant Twins, and the use of the Lawful Insights API™.

WHAT IS FUNCTIONAL SEPARATION?

A report by the European Data Protection Supervisor ("EDPS") – "*Meeting The Challenges of Big Data – A Call For Transparency, User Control, Data Protection By Design And Accountability*" highlighted functional separation as a potential solution for helping to resolve conflicts between innovative data use and data protection.[21]

European Data Protection Supervisor

Opinion 7/2015

Meeting the challenges of big data

A call for transparency, user control, data protection by design and accountability

[20] Controlled Relinkable Data consists of Replacement De-Identifiers (R-DDIDs) and Association De-Identifiers (A-DDIDs), as more fully described herein. The concept of Controlled Linkable Data was presented at an International Association of Privacy Professionals (IAPP) program entitled General Data Protection Regulation (GDPR) Big Data Analytics featuring Gwendal Le Grand, Director of Technology and Innovation at the French Data Protection Authority—the CNIL, Mike Hintze, Partner at Hintze Law and former Chief Privacy Counsel and Assistant General Counsel at Microsoft, and Gary LaFever, CEO and General Counsel at Anonos and former law partner at Hogan Lovells (see https://www.anonos.com/iapp-gdpr-data-analytics-webinar-replay) and explained in a Whitepaper co-authored by Messrs. Hintze and LaFever entitled Meeting Upcoming GDPR Requirements While Maximizing the Full Value of Data Analytics (see https://papers.ssrn.com/sol3/papers. cfm?abstract_id=2927540)

[21] See https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf at page 15.

The principle of functional separation involves using technical and organisational safeguards to separate information value from identity to enable the discovery of trends and correlations independent from applying the insights gained to the data subjects concerned. The EDPS noted at the time of publication of this report that:

> *"There is little evidence of experience with effective implementation of functional separation outside some specialist organisations such as national statistical offices and research institutions. In order to take full advantage of secondary uses of data, it is essential that other organisations develop their expertise and offer comparable guarantees against misuse of data."*[22]

Anonos' eight years of legal and technical research developed the expertise necessary to "offer comparable guarantees against misuse of data" as suggested by the EDPS by leveraging functional separation principles to deliver "Speed To Insight, Lawful & Ethically."

Under the GDPR, the concept of functional separation is embodied in the definitional requirements necessary to achieve and maintain Pseudonymisation as newly defined under Article 4(5). This requires that the information value of data must be separated from the identity of a data subject such that additional securely stored information is necessary to relink information value to identity, and then only for authorised processing under controlled conditions.

The principle of functional separation also exists under other data protection laws using different terms – e.g. heightened "De-Identification" under the California Consumer Privacy Act (CCPA) and the proposed Indian Data Privacy Law, and "Anonymisation" under the Brazilian Data Protection Law.

The CCPA introduces the principle of functional separation through its definition of "Personal Information", which is subject to various protections under the Act. Personal Information includes "information that identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household."

The CCPA's extensive list of data comprising protected Personal Information includes "static" and even "probabilistic" tokens (replacement identifiers) used to replace personal information if "more probable than not" that the information could be used to identify a consumer or device.

While restrictions under the CCPA do not apply to "De-identified Data," traditional approaches to de-identification do not satisfy the heightened requirements for De-identification under the CCPA. CCPA heightened De-identification requirements are not satisfied using "static" and "probabilistic" tokens (replacement identifiers) because they fail to adequately separate information value from identity to prevent unauthorised reidentification of consumers.

LIMITATIONS OF CONSENT

In today's digital-insight-driven world, new technical measures are required to help balance data innovation and protection of individual privacy rights. This is because new approaches to processing information are increasingly moving beyond the capabilities of consent, and for non-consent lawful bases, technical risk-based controls are needed to prove that data is protected.

[22] Id at page 21, footnote 42.

Consent by itself is no longer fair or effective, and its limitations are well articulated in the following quote:

> "*The free and informed consent that today's privacy regime imagines simply cannot be achieved. Collection and processing practices are too complicated. No company can reasonably tell a consumer what is really happening to his or her data. No consumer can reasonably understand it. And if companies can continue to have their way with user data as long as they tell users first, consumers will continue to accept the unacceptable: If they want to reap the benefits of these products, this is the price they will have to pay. Companies should be expected and required to act reasonably to prevent harm to their clients. They should exercise a duty of care. The burden no longer should rest with the user to avoid getting stepped on by a giant. Instead, the giants should have to watch where they're walking.*"[23]

Regulators enforcing the GDPR agree – issued guidance recognises the inadequacy of consent as used in the past. Moreover, the use of consent as a basis for processing is severely restricted, particularly when it comes to iterative secondary uses of personal data such as analytics, machine learning, and AI. New technical safeguards are necessary to support other lawful bases for processing to augment consent to enable lawful and ethical innovation.

PSEUDONYMISATION AS A WAY FORWARD

One of these new technical controls is Pseudonymisation, as newly defined under the GDPR. The ENISA publication *Recommendations on Shaping Technology According to GDPR Provisions: An Overview on Data Pseudonymisation*[24] highlights the following benefits from GDPR-compliant pseudonymisation:

1. Pseudonymisation serves as a vehicle to "relax" certain data controller obligations, including:

   a. Lawful repurposing (further processing) in compliance with purpose limitation principles;
   b. Archiving of data for statistical processing, public interest, scientific or historical research;
   c. Reduced notification obligations in the event of a data breach.

2. Pseudonymisation supports a more favourable (broader) interpretation of data minimisation.

3. Pseudonymisation goes beyond protecting "real-world personal identities" by protecting indirect identifiers.

4. Pseudonymisation provides for unlinkability between data and personal identity, furthering the fundamental data protection principles of necessity and data minimisation.

5. Pseudonymisation decouples privacy and accuracy, enabling Data Protection by Design and by Default while at the same time allowing data about individuals to remain more accurate.

[23] https://www.washingtonpost.com/opinions/our-privacy-regime-is-broken-congress-needs-to-create-new-norms-for-a-digital-age/2019/01/04/c70b228c-0f9d-11e9-8938-5898adc28fa2_story.html

[24] See https://www.anonos.com/hubfs/ENISA_Pseudonymisation_Recomendations_GDPR_November_2018.pdf

While Pseudonymisation has many benefits, using it effectively requires significant expertise. The same ENISA report recognises that effective Pseudonymisation is highly context-dependent and “requires a high level of competence” to prevent attacks while maintaining data utility.

INTRODUCTION TO ANONOS BIGPRIVACY

BigPrivacy’s unique approach:

- Provides a powerful tool for organisations to implement GDPR-mandated Data Protection by Design and by Default.
- Supports a risk-based approach to data protection. It delivers the flexibility and control to provide purpose, context, required utility, and desired scalability of intended processing, while addressing the necessary protection of personal data. This is accomplished by empowering privacy engineers to apply finely tuned combinations of anonymisation techniques, GDPR-compliant Pseudonymisation (and CCPA-compliant heightened de-identification), together with patented risk-based controls.
- Ensures transparency and auditability of privacy-engineering techniques and offers the visibility of the security and data protection levels used to achieve desired accountability.
- Implements ENISA recommendations and best practices for Pseudonymisation.
- Introduces multiple new innovations that advance the state-of-the-art to address new challenges presented by Big Data.

The core of Anonos' capabilities is centered around **Variant Twins**®.

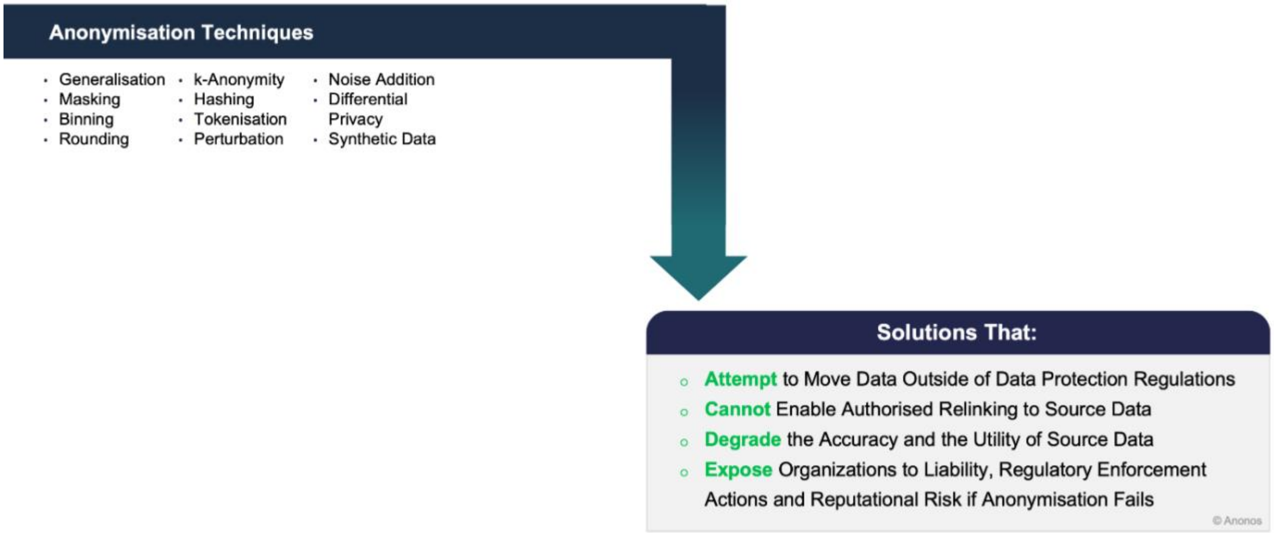
Variant Twins are our patented approach to controlled selective disclosure. With Variant Twins, you **only** provide the level of identifiability needed for each authorised process. Because all Variant Twins are derived from, rather than by permanently altering, the original source data, 100% of the value of source data is retained.

To understand how Variant Twins work, let’s look at how they (i) **leverage** and (ii) **compare** to other privacy enhancing techniques.

TRADITIONAL ANONYMISATION

First, let's consider traditional anonymisation techniques. Anonymisation attempts to remove data from the jurisdiction of data privacy laws. However, if the original data were to be made truly anonymous against all potential risk of reidentification, the anonymised data would lose most of its utility and value.

Figure 1: Shortcomings of Anonymisation

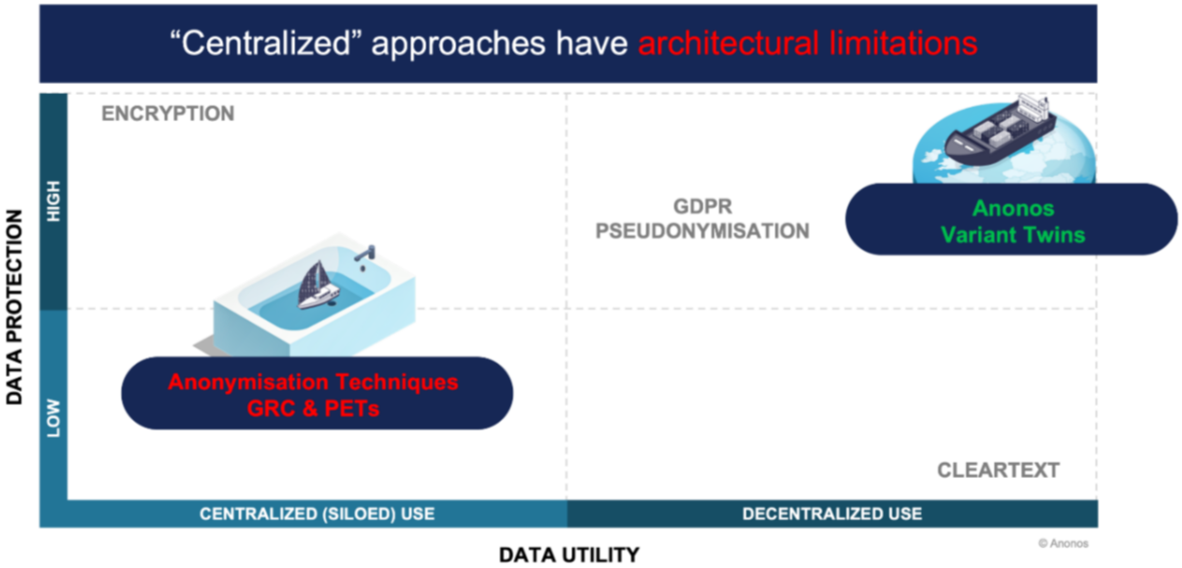

Anonymisation Techniques
Generalisation
Masking
Binning
Rounding
k-Anonymity
Hashing
Tokenisation
Perturbation
Noise Addition
Differential Privacy
Synthetic Data
Solutions That:
Attempt to Move Data Outside of Data Protection Regulations
Cannot Enable Authorised Relinking to Source Data
Degrade the Accuracy and the Utility of Source Data
Expose Organizations to Liability, Regulatory Enforcement Actions and Reputational Risk if Anonymisation Fails
© Anonos

Traditional anonymisation solutions attempt to preserve some level of utility by managing the increased risk of reidentification by restricting processing to enclaves or silos. We refer to this as “Centralised Processing.”

Using this approach means that the data is not available for high value uses such as sharing, combining and enriching because when information is used outside of the centralised processing environment, the risk of unauthorised reidentification via the Mosaic Effect becomes too high. The Mosaic Effect occurs when a person is indirectly identifiable via linkage attacks because the “anonymised” source data can be combined with other pieces of information, enabling the individual to be distinguished from others. More details on this particular shortcoming of anonymisation are available at www.MosaicEffect.com.

Anonymisation techniques, by definition, **cannot** enable authorised relinking; they also degrade the accuracy of data and expose parties in the data supply chain to potential liability.

Figure 2: Choose Data Protection for Your Data Strategy

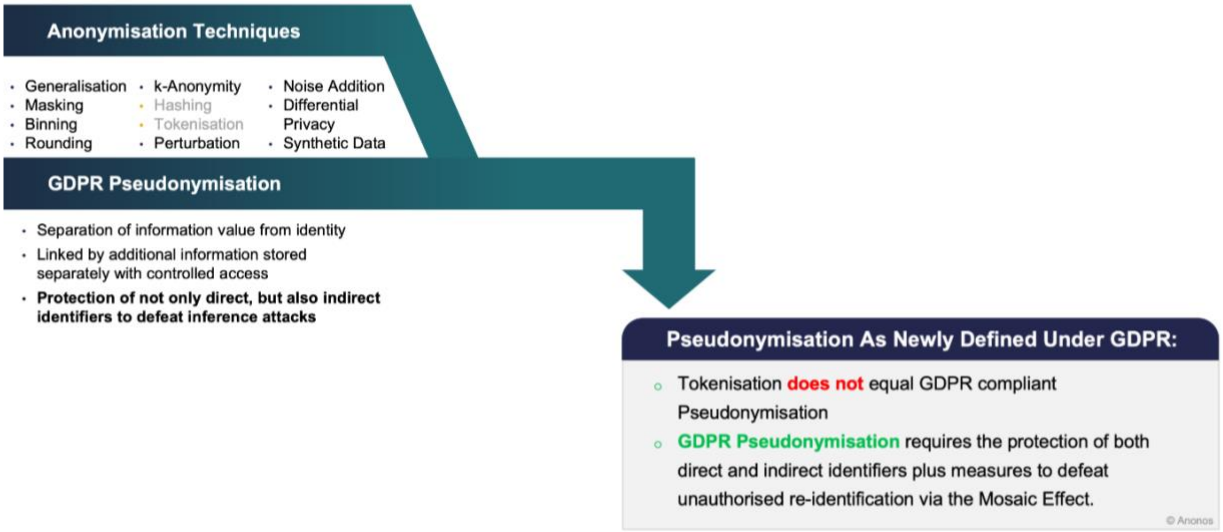


Techniques that protect data for low risk centralised processing do not scale well, because they become **ineffective** in decentralised high value environments like advanced analytics, data sharing, combining and enriching.

Figure 2 above highlights the shortcoming of traditional anonymisation approaches. What works in a centralised environment (depicted by the small boat in a bathtub) simply does not support the "out in the open ocean", high value, decentralised processing necessary for global digitisation.

Traditional anonymisation approaches break down – **they fail to protect data** – when used for decentralised processing. This is because of the real risk of unauthorised reidentification, **which often results in the surveillance of individuals for both commercial and illegal ends.**

Figure 3: GDPR Pseudonymisation Improves Upon Anonymisation



Anonymisation Techniques

- Generalisation
- Masking
- Binning
- Rounding
- k-Anonymity
- Hashing
- Tokenisation
- Perturbation
- Noise Addition
- Differential Privacy
- Synthetic Data

GDPR Pseudonymisation

- Separation of information value from identity
- Linked by additional information stored separately with controlled access
- **Protection of not only direct, but also indirect identifiers to defeat inference attacks**

Pseudonymisation As Newly Defined Under GDPR:

- Tokenisation **does not** equal GDPR compliant Pseudonymisation
- **GDPR Pseudonymisation** requires the protection of both direct and indirect identifiers plus measures to defeat unauthorised re-identification via the Mosaic Effect.

© Anonos

Next let's take a look at what GDPR Pseudonymisation requires and how it improves upon anonymisation.

WHAT IS PSEUDONYMISATION UNDER GDPR?

The first thing to note is that GDPR-compliant Pseudonymisation is an **outcome, not a technique.** This means that old, pre-GDPR approaches (which are too often still incorrectly referred to as "pseudonymisation") will rarely, if ever, meet the GDPR definition of what Pseudonymisation actually requires.

Prior to the GDPR, pseudonymisation was widely understood to mean replacing direct identifiers with tokens. It was a privacy-enhancing technique. **However, Article 4(5) of the GDPR introduces a new legal definition of Pseudonymisation, where it is defined as an outcome.**[25]

In order to satisfy the new requirements for GDPR-compliant Pseudonymisation, you must separate information value from individual identity so that the **only** way to re-identify an individual is by accessing data that is held separately by the data controller.

The second thing to note is that pre-GDPR, pseudonymisation was thought of as a technique **applied to individual fields** within a data set. The new GDPR definition, in combination with the GDPR definition for Personal Data, results in Pseudonymisation being an outcome for the **data set as a whole** (the entire collection of direct identifiers, indirect identifiers and other attributes).

A third observation can be made as a consequence of the massive proliferation of data publicly available for free, privately available for sale, and on the dark web as a result of ongoing daily data breaches globally. It can be best summarised in this quote by Professor Paul Ohm[26]:

> "These results suggest that maybe **everything is PII** to one who has access to the right outside information."[27] *(emphasis added)*

Taken together, the implication is clear: in order to achieve GDPR-compliant Pseudonymisation you have to protect not only direct identifiers, but also indirect identifiers. You must also consider the degree of protection to be applied to all other attributes in a data set while still preserving its utility for the intended use of the data. **Anonos technology does this.**

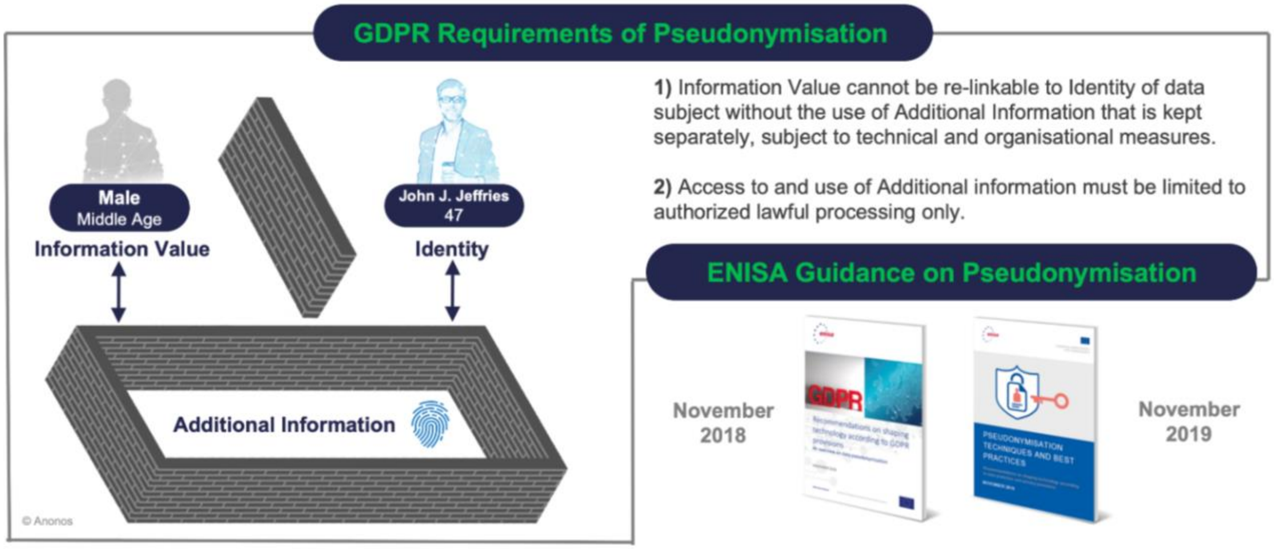
Additional information is available at www.Pseudonymisation.com.

[25] This is why the term Pseudonymisation is used fifteen (15) times in the GDPR, compared Anonymisation which is used only three (3) times, and Encryption which is used only four (4) times in the GDPR. No other Privacy Enhancing Techniques (PETs) are referenced in the GDPR. Benefits of GDPR compliant Pseudonymisation include, but are not limited to, the following: (i) tipping the balance in favour of Legitimate Interests processing (Articles 5(1)(a), 6(1)(f), and WP29 WP 217); (ii) more flexible change of purpose (Article 5(1)(b), WP29 WP 203); (iii) more expansive data minimisation (Articles 5(1)(c), 89(1)); (iv) more flexible storage limitation (Articles 5(1)(e), 89(1)); (v) enhanced security (Articles 5(1)(f), 32); (vi) more expansive further processing (Article 6(4), WP29 WP 217); (vii) more flexible profiling (WP29 WP 251 rev.01 - Annex 1, Recital 71, Article 22); and (viii) ability to lawfully and ethically share, combine and enhance data (recitals 42 and 43, Articles 11(2), 12(2), WP29 WP259 rev.01).

[26] Paul Ohm is a Professor of Law at the Georgetown University Law Center on Privacy and Technology in Washington DC.

[27] https://www.uclalawreview.org/broken-promises-of-privacy-responding-to-the-surprising-failure-of-anonymization-2/ p.1723

Figure 4: GDPR Pseudonymisation is an Outcome, Not a Technique


GDPR Requirements of Pseudonymisation
Male
Middle Age
Information Value
John J. Jeffries
47
Identity
Additional Information
© Anonos
ENISA Guidance on Pseudonymisation
November 2018
November 2019
GDPR
PSEUDONYMISATION TECHNIQUES AND BEST PRACTICES

1) Information Value cannot be re-linkable to Identity of data subject without the use of Additional Information that is kept separately, subject to technical and organisational measures.

2) Access to and use of Additional information must be limited to authorized lawful processing only.

So, what does that mean?

It means in nearly all cases, organisations and businesses need new technology to implement Pseudonymisation because it is no longer a technique, but an **outcome.** It requires that the only way you can get back and forth over the wall shown at the top left of Figure 4 above (between "information value" and "identity") is via access to additional information that is kept separately by the data controller.

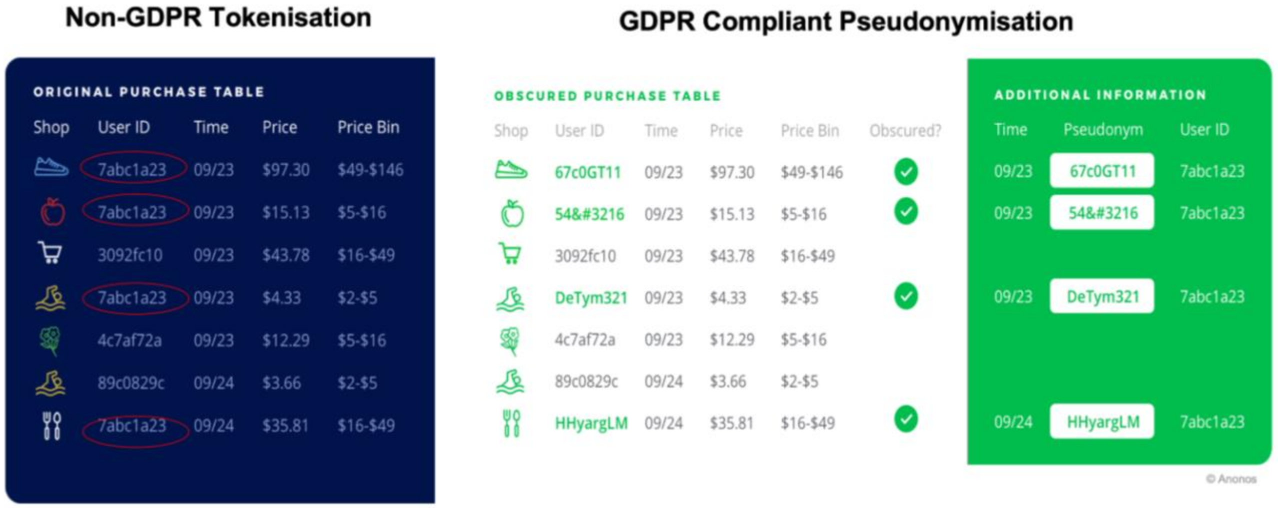
**If you can re-link data without access to this separately held additional information, it is *not* GDPR-compliant Pseudonymisation.
It also means that the data was not successfully anonymised.**

The extent and specificity of the technical requirements necessary for achieving GDPR-compliant Pseudonymisation are significantly underappreciated. ENISA has outlined over 50 requirements for implementing GDPR Pseudonymisation. One of them is tokenisation – **but that's essentially all that any vendor other than Anonos does.**

Additional information is available at www.ENISAguidelines.com.[28]

[28] See Appendix B for Cross Reference to ENISA Pseudonymisation Guidance.

Figure 5: GDPR Pseudonymisation Counters Mosaic Effect



Non-GDPR Tokenisation

ORIGINAL PURCHASE TABLE

Shop	User ID	Time	Price	Price Bin
👟	7abc1a23	09/23	$97.30	$49-$146
🍎	7abc1a23	09/23	$15.13	$5-$16
🛒	3092fc10	09/23	$43.78	$16-$49
🏊	7abc1a23	09/23	$4.33	$2-$5
🌹	4c7af72a	09/23	$12.29	$5-$16
🏊	89c0829c	09/24	$3.66	$2-$5
🍴	7abc1a23	09/24	$35.81	$16-$49

GDPR Compliant Pseudonymisation

OBSCURED PURCHASE TABLE

Shop	User ID	Time	Price	Price Bin	Obscured?
👟	67c0GT11	09/23	$97.30	$49-$146	✓
🍎	54&#3216	09/23	$15.13	$5-$16	✓
🛒	3092fc10	09/23	$43.78	$16-$49	
🏊	DeTym321	09/23	$4.33	$2-$5	✓
🌹	4c7af72a	09/23	$12.29	$5-$16	
🏊	89c0829c	09/24	$3.66	$2-$5	
🍴	HHyargLM	09/24	$35.81	$16-$49	✓

ADDITIONAL INFORMATION

Time	Pseudonym	User ID
09/23	67c0GT11	7abc1a23
09/23	54&#3216	7abc1a23
09/23	DeTym321	7abc1a23
09/24	HHyargLM	7abc1a23

© Anonos

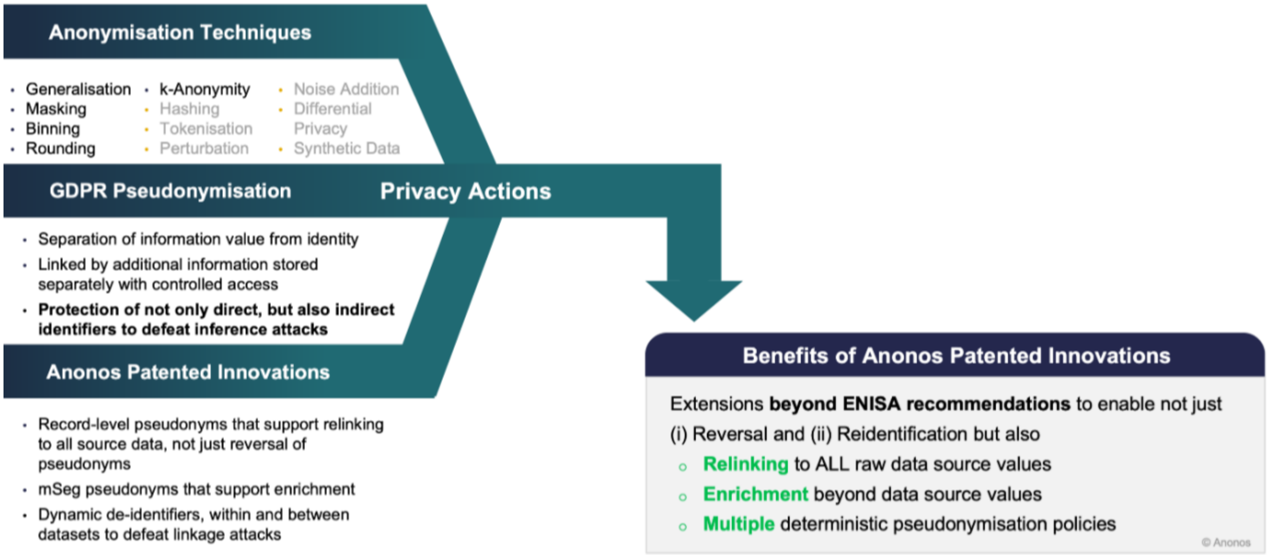
Figure 5 above, which provides a summary of information available at www.MosaicEffect.com, shows an example of what people mean when referring to pseudonymisation in its pre-GDPR form as a **technique**. Here a username is replaced with a token in the form of a User ID, but the same token is used repeatedly for each occurrence of the same User ID. This is called **static** (or persistent) tokenisation.

So, what do you need in order to satisfy the GDPR **outcome** requirement for Pseudonymisation? You must have **dynamism** in the way that you allocate and change tokens.

Figure 5 shows that in each place that the static token 7abc1a23 was previously used, when applying dynamism, it is replaced with a **different** pseudonym each time. This means that the only way to get to the identity of the individual represented by the User ID 7abc1a23 is by accessing separately kept "additional information."

THE ADVANTAGES OF CONTROLLED LINKABLE DATA

Figure 6: Anonos Controlled Linkable Data®



Anonymisation Techniques

- Generalisation
- Masking
- Binning
- Rounding
- k-Anonymity
- Hashing
- Tokenisation
- Perturbation
- Noise Addition
- Differential Privacy
- Synthetic Data

GDPR Pseudonymisation

Privacy Actions

- Separation of information value from identity
- Linked by additional information stored separately with controlled access
- **Protection of not only direct, but also indirect identifiers to defeat inference attacks**

Anonos Patented Innovations

- Record-level pseudonyms that support relinking to all source data, not just reversal of pseudonyms
- mSeg pseudonyms that support enrichment
- Dynamic de-identifiers, within and between datasets to defeat linkage attacks

Benefits of Anonos Patented Innovations

Extensions **beyond ENISA recommendations** to enable not just (i) Reversal and (ii) Reidentification but also

- **Relinking** to ALL raw data source values
- **Enrichment** beyond data source values
- **Multiple** deterministic pseudonymisation policies

© Anonos

Having looked at anonymisation and GDPR-compliant Pseudonymisation, let's now look at Anonos' patented innovations that go beyond ENISA requirements to enable organisations to "take full advantage of secondary uses of data" by delivering "comparable guarantees against misuse of data" by leveraging functional separation as suggested by the EDPS.[29]

Anonos enables you to do **more** than what GDPR-compliant Pseudonymisation requires. Anonos enables the reversal of pseudonymous tokens and authorised reidentification, both of which are powerful benefits of GDPR-compliant Pseudonymisation.

However, Anonos goes further by enabling a data controller (for authorised purposes and under controlled conditions) to relink to **any or all values from the source data**. This is done using Anonos patented Controlled Linkable Data[30], which represents a significant advance over GDPR compliant Pseudonymisation.

As noted above, pseudonymisation when practiced in the pre-GDPR era as a technique has been one-dimensional: static tokens applied to direct identifiers, where a specific identifier is assigned the same token consistently both within and between databases. ENISA refers to this as a "pseudonymisation policy" and refers to these static tokens as being fully deterministic, meaning that a consistent input will result in a deterministic or consistent token. While useful as a localised security

[29] See https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf at page 42.

[30] Controlled Linkable Data consists of Replacement De-Identifiers (R-DDIDs) and Association De-Identifiers (A-DDIDs), as more fully described in herein. The concept of Controlled Linkable Data was presented at an International Association of Privacy Professionals (IAPP) program entitled General Data Protection Regulation (GDPR) Big Data Analytics featuring Gwendal Le Grand, Director of Technology and Innovation at the French Data Protection Authority—the CNIL, Mike Hintze, Partner at Hintze Law and former Chief Privacy Counsel and Assistant General Counsel at Microsoft, and Gary LaFever, CEO and General Counsel at Anonos and former law partner at Hogan Lovells (see https://www.anonos.com/iapp-gdpr-data-analytics-webinar-replay) and explained in a Whitepaper co-authored by Messrs. Hintze and LaFever entitled Meeting Upcoming GDPR Requirements While Maximizing the Full Value of Data Analytics (see https://papers.ssrn.com/sol3/papers. cfm?abstract_id=2927540)

technique, this approach provides limited protection against unauthorised reidentification because it is so vulnerable to linkage attacks and inference attacks.

An additional pseudonymisation policy described by ENISA in their guidance at the other end of the spectrum is fully randomised pseudonymisation: a specific identifier receives a different pseudonym every time it occurs. This maximises protection but significantly reduces data utility.

Anonos Controlled Relinkable Data provides improvements over newly defined Pseudonymization under the GDPR that advance the state-of-the-art in at least three important ways:

1. Granular control over relinking to source data by using record-level identifiers, unlike mere reversal of pseudonymous tokens associated with traditional techniques. This makes it possible to retrieve additional information from the source data beyond the data included in an original Variant Twin.
2. Powerful resistance to reidentification by adversaries attempting linkage and inference attacks, while preserving much higher levels of analytical utility than previously obtainable. This is done by enabling efficient application of deterministic pseudonymisation to individual indirect identifiers, or combinations of them, and not just to direct identifiers.
3. Support for not only ENISA-defined fully-randomised and fully deterministic pseudonymisation policies, but also for three additional intermediate pseudonymisation policies. Drawing on the ENISA nomenclature, these can be characterised as field, table, and document deterministic pseudonymisation respectively.
   - **Field Deterministic Pseudonymisation:** consistency is maintained only within individual columns in a table, with different pseudonyms being used between columns containing the same data (e.g. country of origin, and country of residence within a single table) as well as columns in other tables/databases. This is the default for BigPrivacy. This ensures that each occurrence of the same data value in a field is replaced by the same pseudonym within that column. The same data values in other columns would be replaced by a different set of pseudonyms.
   - **Table Deterministic Pseudonymisation:** consistency is maintained within a table. Multiple (or all) fields within single table or data set have pseudonym values that are deterministic within that table/data set, but different pseudonyms are used in each succeeding table/data set.
   - **Document Deterministic Pseudonymisation:** all occurrences of a data value within all fields in all tables in one database are assigned the same pseudonym. New pseudonyms are used for occurrences of that data value in each succeeding database.

BigPrivacy supports this range of pseudonymisation policies by leveraging three different ENISA-recommended cryptography techniques as pseudonymisation functions, each of which are characterised by ENISA as providing strong data protection:

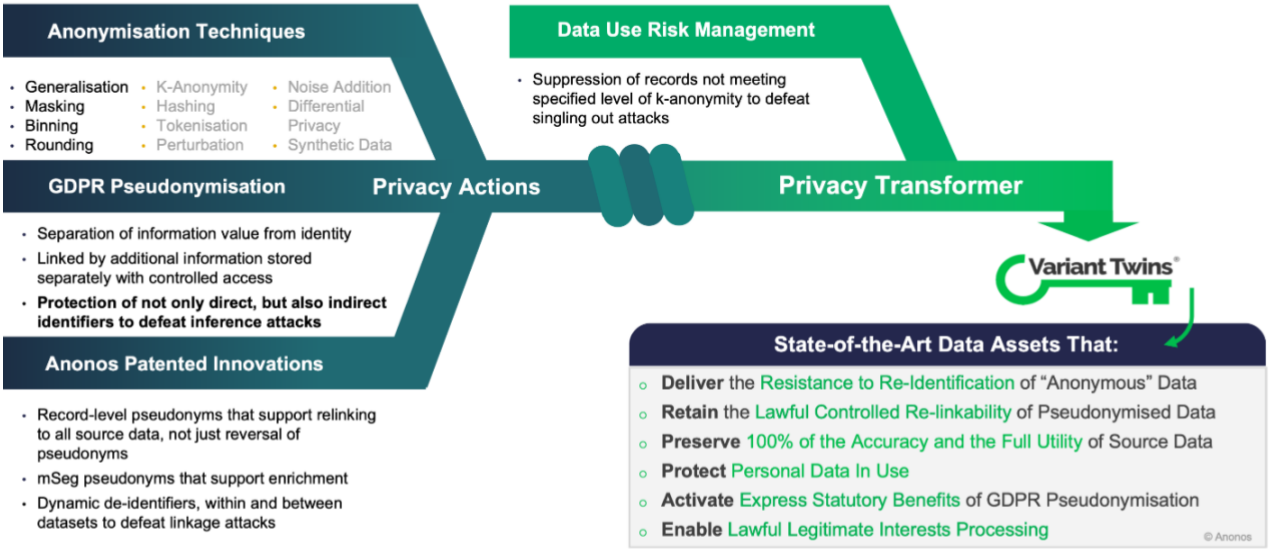
- **Cryptographic Pseudo-Random Number Generation (CRNG):** BigPrivacy leverages the computer operating system entropy pool to create Replacement Dynamic De-Identifiers (R-DDIDs®) for individual field values for fully randomised pseudonymisation.
- **Hashed Message Authentication Code (HMAC):** HMAC is used to create non-reversible (but still re-linkable) Association Dynamic De-Identifiers (A-DDIDs®) for four of the deterministic pseudonymisation policies mentioned above field, table, document, and fully deterministic).
- **Symmetric AES Encryption:** BigPrivacy uses symmetric AES encryption to fully randomise R-DDIDs used as record level pseudonyms and for two types of A-DDIDs: reversible deterministic and reversable fully randomised. These are useful in circumstances where creating a master index (mapping table) is not desired and/or where pseudonym reversal without relinking is useful (e.g., interpreting results from Machine Learning and AI models created using Variant Twins).

For the first two techniques, a recovery function is provided via a securely and separately stored mapping table. This enables reversal of Pseudonymisation when authorised. The mapping table is the “additional information” kept separately subject to technical and organisational measures to ensure that the personal data are not attributable to an identified or identifiable natural person under the GDPR definition of Pseudonymisation. For the third technique a mapping table is not required due to the inherent reversibility of symmetric encryption, with the keys necessary for decryption being the “information held separately”.

K-ANONYMITY AND BIGPRIVACY VARIANT TWINS

As shown in Figure 7below, the final process step in creating Variant Twins involves leveraging k-anonymity to enable Data Use Risk Management. This provides protection against reidentification attacks using singling out.

Figure 7: Protect Data in Use, for Speed To Insight, Lawfully & Ethically


Anonymisation Techniques
Generalisation
Masking
Binning
Rounding
K-Anonymity
Hashing
Tokenisation
Perturbation
Noise Addition
Differential Privacy
Synthetic Data
Data Use Risk Management
Suppression of records not meeting specified level of k-anonymity to defeat singling out attacks
GDPR Pseudonymisation
Privacy Actions
Privacy Transformer
Separation of information value from identity
Linked by additional information stored separately with controlled access
Protection of not only direct, but also indirect identifiers to defeat inference attacks
Anonos Patented Innovations
Record-level pseudonyms that support relinking to all source data, not just reversal of pseudonyms
mSeg pseudonyms that support enrichment
Dynamic de-identifiers, within and between datasets to defeat linkage attacks
Variant Twins®
State-of-the-Art Data Assets That:
Deliver the Resistance to Re-Identification of “Anonymous” Data
Retain the Lawful Controlled Re-linkability of Pseudonymised Data
Preserve 100% of the Accuracy and the Full Utility of Source Data
Protect Personal Data In Use
Activate Express Statutory Benefits of GDPR Pseudonymisation
Enable Lawful Legitimate Interests Processing
© Anonos

The following overview of k-anonymity is derived from a description provided by the U.S. Department of Health & Human Services (HHS).[31]

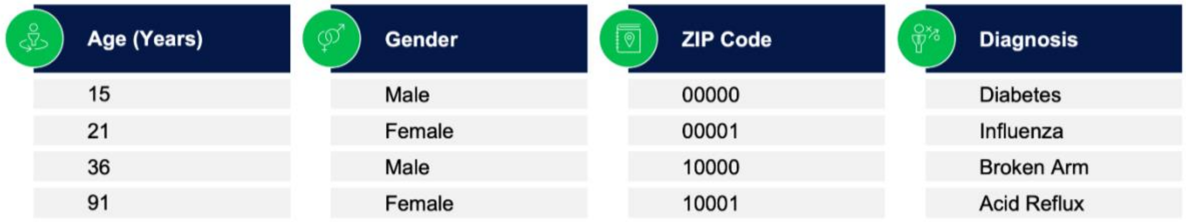
> When using the k-anonymity technique, “k” refers to the number of people to which each disclosed record might correspond. In practice, this correspondence is assessed using the features that could be reasonably applied by a recipient to identify an individual data subject.

Table 2 below illustrates an application of generalisation and suppression methods to achieve a k-anonymity value of “2” (2-anonymity) with respect to the Age, Gender, and ZIP Code columns in the fictitious protected health information included in Table 1 below. All rows correspond to fictitious

[31] https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

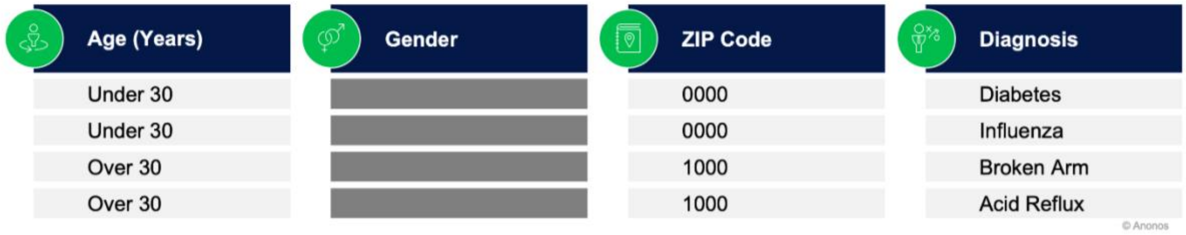
patient records with the same combination of generalised and suppressed values for Age, Gender, and ZIP Code. Notice that Gender has been suppressed completely.

Table 1
Protected Health Information (PHI)



Age (Years)	Gender	ZIP Code	Diagnosis
15	Male	00000	Diabetes
21	Female	00001	Influenza
36	Male	10000	Broken Arm
91	Female	10001	Acid Reflux

Table 2
K-Anonymity Level of "2"



Age (Years)	Gender	ZIP Code	Diagnosis
Under 30		0000	Diabetes
Under 30		0000	Influenza
Over 30		1000	Broken Arm
Over 30		1000	Acid Reflux

© Anonos

The combination of:

1. Anonymisation techniques;
2. Pseudonymisation/heightened de-identification techniques; and
3. Anonos patented Controlled Linkable Data risk-based controls;

used for a particular field in a dataset is what we call a "Privacy Action."™

The combination of different Privacy Actions used for a data set, together with the selected level of k-anonymity, comprise what we call a "Privacy Transformer."™ When a source data set is run through a Privacy Transformer, the result is a Variant Twin. A Variant Twin is a version of the source data transformed by the selected Privacy Actions and filtered for reidentification risk to suppress records that do not meet the required k-anonymity threshold.

This combination of Privacy Actions and reidentification risk management provides tailored protection against:

- Unauthorised combining of data with other data sources; which can result in
- Unauthorised Re-identification of data subjects; while
- Preserving full data utility that enables compliant secondary uses of data for analytics, AI, and marketing.

The flexibility of this approach enables a privacy engineer to create Variant Twins for different contexts, uses, and risks. This flexibility opens a range of levels of risk-based data protection, from "local protection" for use within a locally controlled enclave or siloed environment to "global protection", enabling lawful and ethical decentralised data sharing, combining and enrichment.

In summary, Variant Twins:

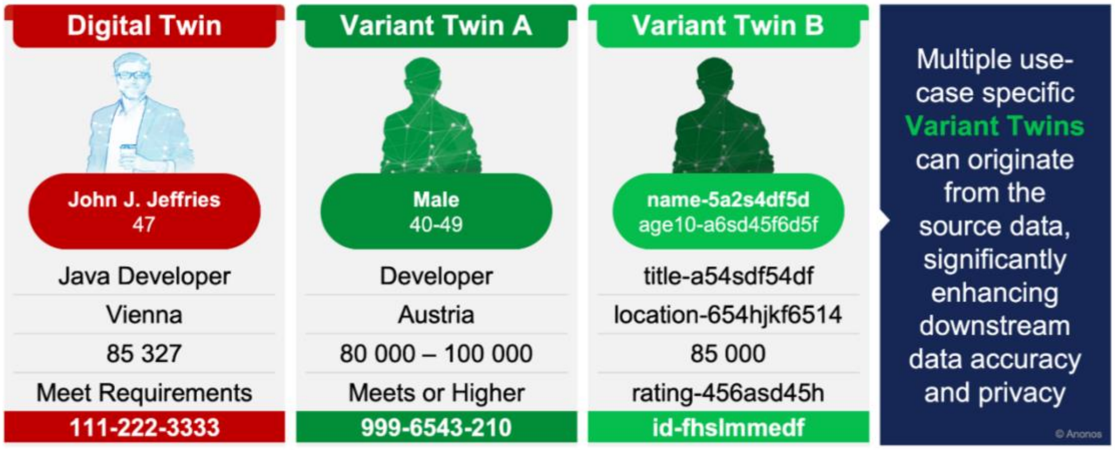
- Deliver resistance to reidentification of truly anonymous data without forcing a data controller to defend the difficult status of "Anonymous" data under the GDPR, by delivering GDPR compliant Pseudonymised data instead.[32]
- Enable data controllers to enforce risk-based control over the re-linkability of data.
- Preserve 100% of the utility of source data.
- Protect data in use.
- Activate express statutory benefits.
- Enable processing under the lawful basis of Legitimate Interests.

Anonos enables organisations to accelerate speed to insight, lawfully and ethically for innovative uses of data.

ENABLING DIGITAL INSIGHTS WITH VARIANT TWINS

Figure 8 below shows a simplified example of data elements that Variant Twins can include.

Figure 8: Digital Twins



Digital Twin	Variant Twin A	Variant Twin B
John J. Jeffries 47	Male 40-49	name-5a2s4df5d age10-a6sd45f6d5f
Java Developer	Developer	title-a54sdf54df
Vienna	Austria	location-654hjkf6514
85 327	80 000 – 100 000	85 000
Meet Requirements	Meets or Higher	rating-456asd45h
111-222-3333	999-6543-210	id-fhslmmedf

Multiple use-case specific Variant Twins can originate from the source data, significantly enhancing downstream data accuracy and privacy

© Anonos

The left-hand side shows a Digital Twin of "John J Jeffries" – a digital representation of a specific person.[33] This includes direct identifiers like name and location, as well as indirect identifiers like date of birth, zip code, income and loan details. In this example, this Digital Twin is the original source data.

[32] A data controller cannot in good faith claim the benefits of "Anonymisation" when decentralised processing makes it impossible to be aware "…of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly" as required under GDPR Recital 26. Claiming "Anonymisation" exposes a data controller to potential liability if re-linkability and unauthorised reidentification is possible.

[33] A digital twin is a digital replica of a living or non-living physical entity. The term refers to a digital replica of potential and actual physical assets (physical twin), processes, people, places, systems and devices that can be used for various purposes. See https://en.wikipedia.org/wiki/Digital_twin

By selecting different Privacy Actions, different versions of Variant Twins can be created to support different use cases and purposes. For example, you might choose to reveal a town or city instead of an exact address or a binned age, income or loan range instead of identifying amounts. This decision would be based on the risks associated with the desired processing of the data.

Variant Twin A has had only limited protections applied, primarily generalisation, along with a format-preserving pseudonym for the ID number. Variant Twin B on the other hand has been much more aggressively transformed, with almost all fields pseudonymised. Note that age, title, location and rating are likely represented by deterministic pseudonyms and so this Variant Twin still retains significant analytic value.

Centralised approaches to data protection approach privacy and data utility as irreconcilable objectives. Anonos decentralised data protection simultaneously enables both goals by enforcing use-case specific risk-based protections embedded in Variant Twins.

Variant Twins enable privacy engineers to achieve the benefits of the first goal of data protection while also achieving the second goal of maximising data utility. This allows organisations to have it both ways i.e. enabling them to *Have Their Cake and Eat it Too*.™

We will now walk through how to create Anonos Variant Twins to enable **Speed To Insight, Lawfully & Ethically.**

Figure 9: Transform Clear Text into Variant Twin

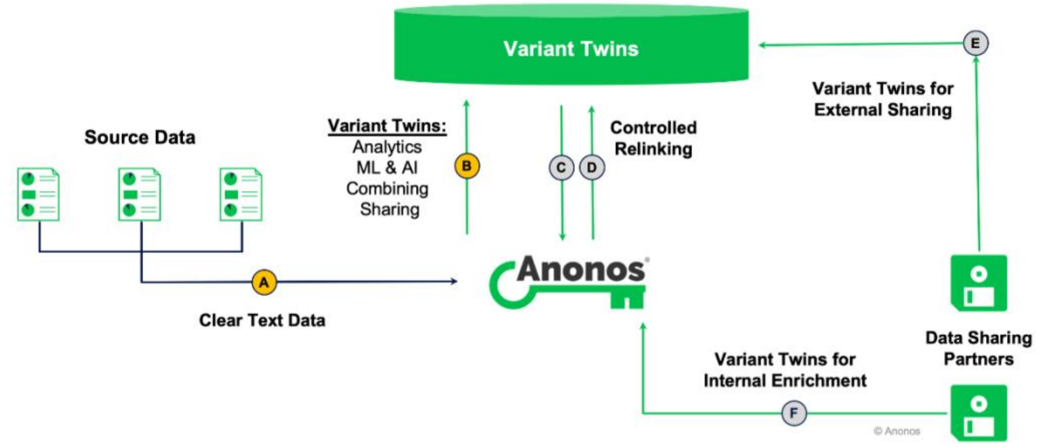
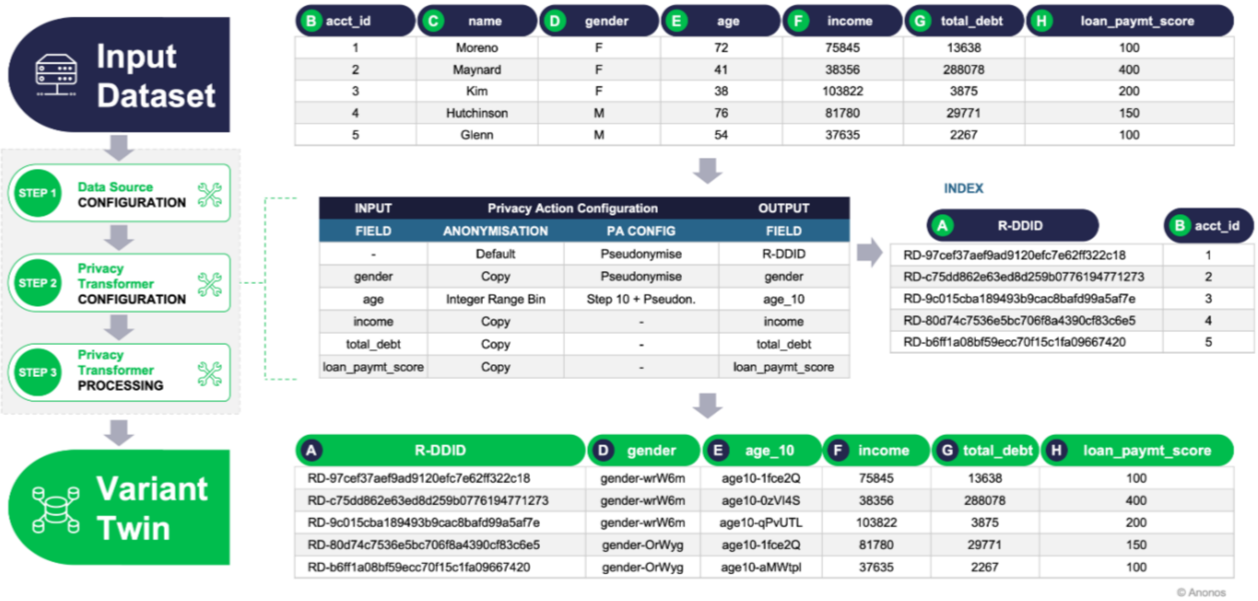


Figure 9 above illustrates how Anonos Technology works to transform clear text data into a Variant Twin.

Upon ingestion, an R-DDID (Random pseudonym) is created for each record and pre-pended to it. Then the R-DDID and source data are written to a master index to allow for later use in relinking when authorised. The R-DDID itself is only a pointer to the original record, and contains no information value itself, so it poses no risk when it is kept with the Variant Twin data because it is only re-linkable by an authorised individual with access to the “additional information” held separately.

Figure 10: Privacy Actions > Privacy Transformer > Variant Twin


Input Dataset
STEP 1 Data Source CONFIGURATION
STEP 2 Privacy Transformer CONFIGURATION
STEP 3 Privacy Transformer PROCESSING
Variant Twin

B acct_id	C name	D gender	E age	F income	G total_debt	H loan_paymt_score
1	Moreno	F	72	75845	13638	100
2	Maynard	F	41	38356	288078	400
3	Kim	F	38	103822	3875	200
4	Hutchinson	M	76	81780	29771	150
5	Glenn	M	54	37635	2267	100

INPUT	Privacy Action Configuration		OUTPUT
FIELD	ANONYMISATION	PA CONFIG	FIELD
-	Default	Pseudonymise	R-DDID
gender	Copy	Pseudonymise	gender
age	Integer Range Bin	Step 10 + Pseudon.	age_10
income	Copy	-	income
total_debt	Copy	-	total_debt
loan_paymt_score	Copy	-	loan_paymt_score

INDEX

A R-DDID	B acct_id
RD-97cef37aef9ad9120efc7e62ff322c18	1
RD-c75dd862e63ed8d259b0776194771273	2
RD-9c015cba189493b9cac8bafd99a5af7e	3
RD-80d74c7536e5bc706f8a4390cf83c6e5	4
RD-b6ff1a08bf59ecc70f15c1fa09667420	5

A R-DDID	D gender	E age_10	F income	G total_debt	H loan_paymt_score
RD-97cef37aef9ad9120efc7e62ff322c18	gender-wrW6m	age10-1fce2Q	75845	13638	100
RD-c75dd862e63ed8d259b0776194771273	gender-wrW6m	age10-0zVI4S	38356	288078	400
RD-9c015cba189493b9cac8bafd99a5af7e	gender-wrW6m	age10-qPvUTL	103822	3875	200
RD-80d74c7536e5bc706f8a4390cf83c6e5	gender-OrWyg	age10-1fce2Q	81780	29771	150
RD-b6ff1a08bf59ecc70f15c1fa09667420	gender-OrWyg	age10-aMWtpl	37635	2267	100

© Anonos

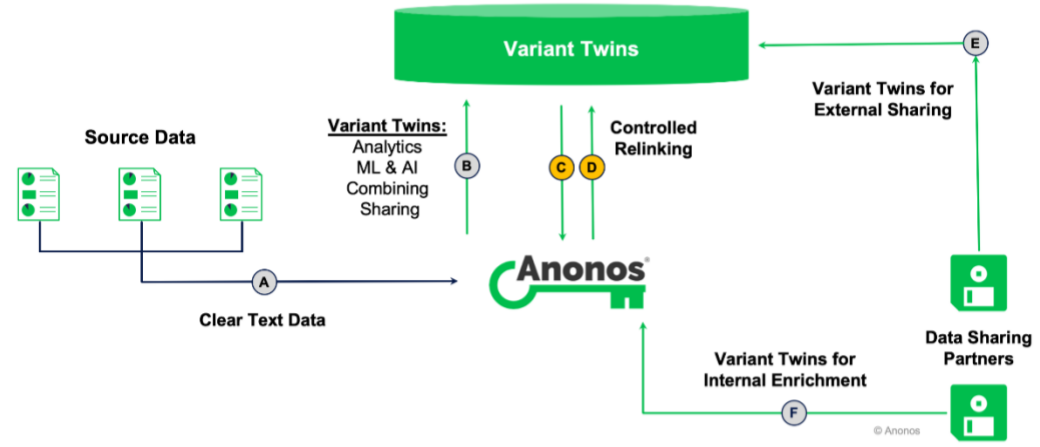
The values indicated at the bottom of Figure 10 above in columns A, D, E, F, G, and H indicate data columns that are a part of the Variant Twin.

The **indirect identifiers** included in this Variant Twin are D and E – "gender" and "age" whereas F, G, and H – representing "income," "total debt," and "loan payment score" – are **attributes**. The **direct identifiers** (sometimes referred to using the legacy term PII) in columns B and C – "acct_id" and "name" – are omitted from the Variant Twin entirely.

By using field deterministic pseudonyms to represent indirect identifiers such as age ranges and gender, data analysts can process data without knowledge of the actual values, thus allowing the analysis to be more privacy-respectful and less identifying but without sacrificing utility.

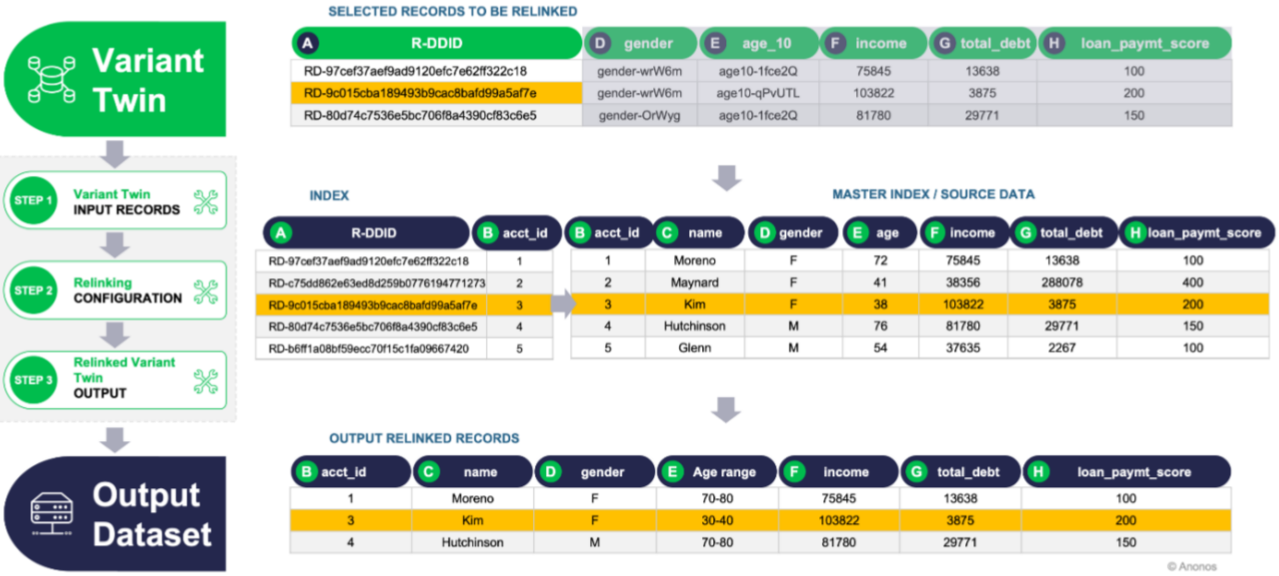
This also reduces the risk of conscious or subconscious bias since data analysts cannot see the values underlying the pseudonyms and thus cannot make assumptions about the data subjects.

Figure 11: Controlled Relinking



Having ingested raw data and created a Variant Twin, we will now walk through controlled relinking (represented by "C" and "D" in Figure 11 above).

Figure 12: Variant Twin Relinking from Source Data



SELECTED RECORDS TO BE RELINKED

A R-DDID	D gender	E age_10	F income	G total_debt	H loan_paymt_score
RD-97cef37aef9ad9120efc7e62ff322c18	gender-wrW6m	age10-1fce2Q	75845	13638	100
RD-9c015cba189493b9cac8bafd99a5af7e	gender-wrW6m	age10-qPvUTL	103822	3875	200
RD-80d74c7536e5bc706f8a4390cf83c6e5	gender-OrWyg	age10-1fce2Q	81780	29771	150

INDEX

A R-DDID	B acct_id
RD-97cef37aef9ad9120efc7e62ff322c18	1
RD-c75dd862e63ed8d259b0776194771273	2
RD-9c015cba189493b9cac8bafd99a5af7e	3
RD-80d74c7536e5bc706f8a4390cf83c6e5	4
RD-b6ff1a08bf59ecc70f15c1fa09667420	5

MASTER INDEX / SOURCE DATA

B acct_id	C name	D gender	E age	F income	G total_debt	H loan_paymt_score
1	Moreno	F	72	75845	13638	100
2	Maynard	F	41	38356	288078	400
3	Kim	F	38	103822	3875	200
4	Hutchinson	M	76	81780	29771	150
5	Glenn	M	54	37635	2267	100

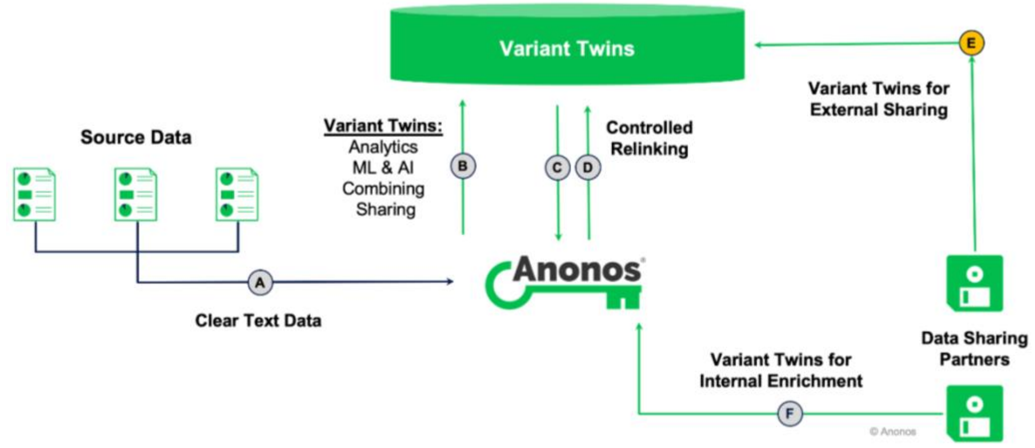
OUTPUT RELINKED RECORDS

B acct_id	C name	D gender	E Age range	F income	G total_debt	H loan_paymt_score
1	Moreno	F	70-80	75845	13638	100
3	Kim	F	30-40	103822	3875	200
4	Hutchinson	M	70-80	81780	29771	150

In Figure 12 above, we show how a data controller is able to relink from a Variant Twin back to the source data when authorised.

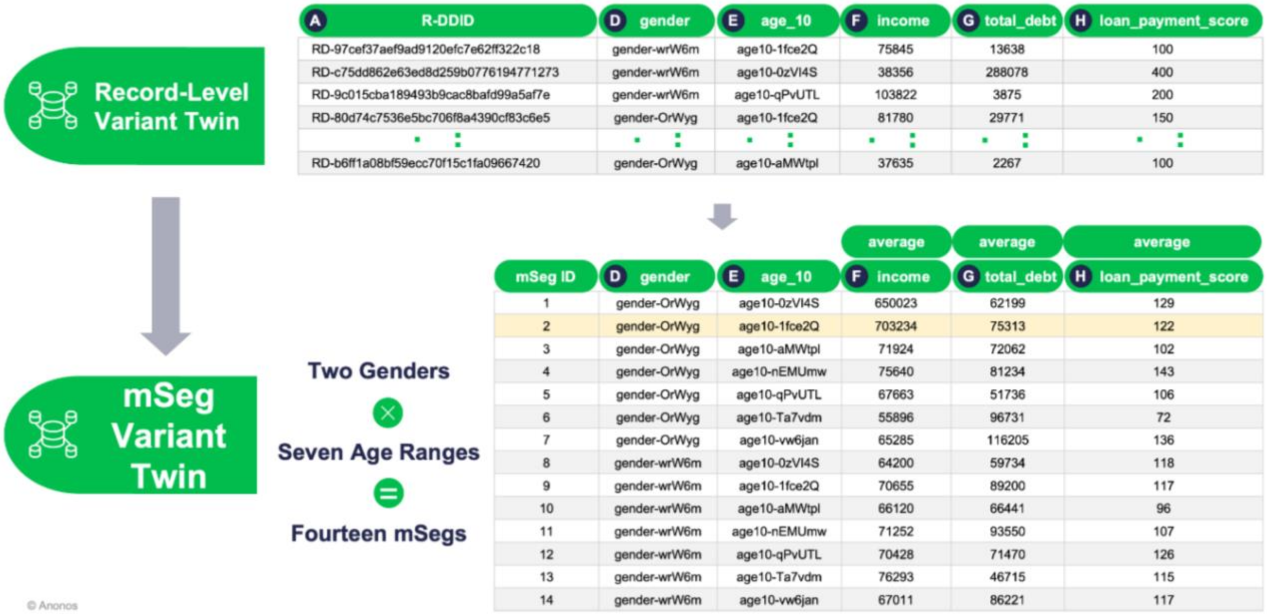
At the top of the figure you can see the Variant Twin, and at the bottom you can see the relinked data from the original raw data set. The row replacement Pseudonyms (R-DDIDs) at the top enable you to relink to **any or all of the original data values** whether or not those values are included in the Variant Twin. This is because each R-DDID serves as a pointer to the entirety of the associated original record via the master index. This patented capability enables multiple layers of abstraction (and privacy) while enabling a data controller to traverse between the layers to gain access to any original source data value for authorised processing. In this way, 100% of the utility of the source data is preserved.

Figure 13: External Sharing



Next, let's look at how a group called a "Microsegment" (or "mSeg") can be created using Variant Twins to enable privacy-respectful data sharing ("E" in Figure 13 above).

Figure 14: Record-level Variant Twin > mSeg Variant Twin



A R-DDID	D gender	E age_10	F income	G total_debt	H loan_payment_score
RD-97cef37aef9ad9120efc7e62ff322c18	gender-wrW6m	age10-1fce2Q	75845	13638	100
RD-c75dd862e63ed8d259b0776194771273	gender-wrW6m	age10-0zVI4S	38356	288078	400
RD-9c015cba189493b9cac8bafd99a5af7e	gender-wrW6m	age10-qPvUTL	103822	3875	200
RD-80d74c7536e5bc706f8a4390cf83c6e5	gender-OrWyg	age10-1fce2Q	81780	29771	150
. :	. :	. :	. :	. :	. :
RD-b6ff1a08bf59ecc70f15c1fa09667420	gender-OrWyg	age10-aMWtpl	37635	2267	100

			average	average	average
mSeg ID	D gender	E age_10	F income	G total_debt	H loan_payment_score
1	gender-OrWyg	age10-0zVI4S	650023	62199	129
2	gender-OrWyg	age10-1fce2Q	703234	75313	122
3	gender-OrWyg	age10-aMWtpl	71924	72062	102
4	gender-OrWyg	age10-nEMUmw	75640	81234	143
5	gender-OrWyg	age10-qPvUTL	67663	51736	106
6	gender-OrWyg	age10-Ta7vdm	55896	96731	72
7	gender-OrWyg	age10-vw6jan	65285	116205	136
8	gender-wrW6m	age10-0zVI4S	64200	59734	118
9	gender-wrW6m	age10-1fce2Q	70655	89200	117
10	gender-wrW6m	age10-aMWtpl	66120	66441	96
11	gender-wrW6m	age10-nEMUmw	71252	93550	107
12	gender-wrW6m	age10-qPvUTL	70428	71470	126
13	gender-wrW6m	age10-Ta7vdm	76293	46715	115
14	gender-wrW6m	age10-vw6jan	67011	86221	117

A useful way to think of mSegs is as **look-alike audiences** that are small enough to represent the distinct characteristics, attributes, preferences, activities, behaviours and even location of a real group of data subjects (all of which may be necessary to achieve business objectives from processing data), but large enough that they don't enable singling out, linking to, or inferences about the identity of individual data subjects.

Each record in a Variant Twin represents a single individual. By comparison, each record in the mSeg represents a small group of individuals that share the same characteristics.

In the example in Figure 14 above, we are using two gender and seven age range field deterministic pseudonyms (A-DDID's) to create microsegments (mSegs). Aggregating the original Variant Twin records by each of the combinations of the two gender pseudonyms and the seven age range pseudonyms results in fourteen mSegs.

This approach can be extended to as many segmentation variables as the data set sample size will support to satisfy the particular mSeg requirements for a specific use case.

Figure 15: Internal Enrichment

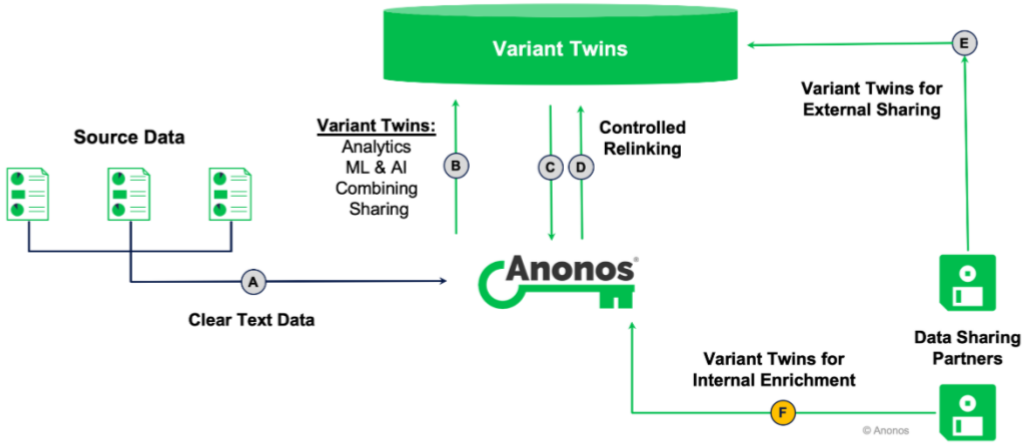
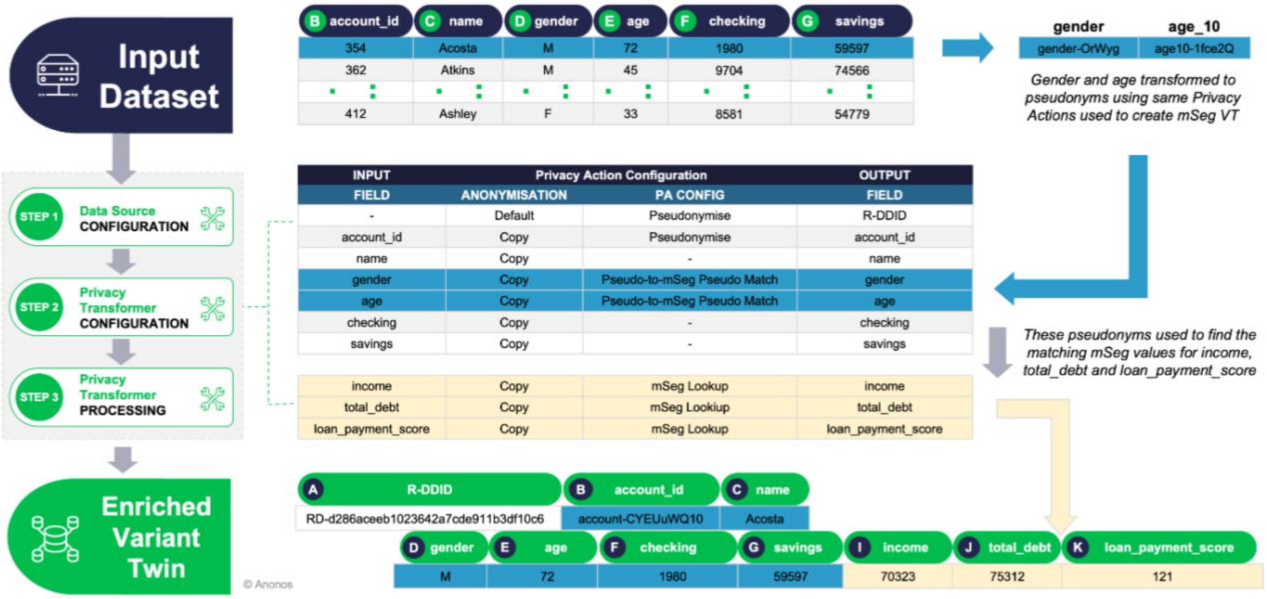


Figure 16 below illustrates how mSegs can be used to enable privacy-respectful data sharing between and among data stewards for an improved customised experience for customers while protecting their privacy (represented by "F" in Figure 15 above).

Figure 16: MSeg Enabled Data Sharing for Enrichment


Input Dataset
STEP 1 Data Source CONFIGURATION
STEP 2 Privacy Transformer CONFIGURATION
STEP 3 Privacy Transformer PROCESSING
Enriched Variant Twin
© Anonos

B account_id	C name	D gender	E age	F checking	G savings
354	Acosta	M	72	1980	59597
362	Atkins	M	45	9704	74566
⋮	⋮	⋮	⋮	⋮	⋮
412	Ashley	F	33	8581	54779

gender	age_10
gender-OrWyg	age10-1fce2Q

Gender and age transformed to pseudonyms using same Privacy Actions used to create mSeg VT

INPUT	Privacy Action Configuration		OUTPUT
FIELD	ANONYMISATION	PA CONFIG	FIELD
-	Default	Pseudonymise	R-DDID
account_id	Copy	Pseudonymise	account_id
name	Copy	-	name
gender	Copy	Pseudo-to-mSeg Pseudo Match	gender
age	Copy	Pseudo-to-mSeg Pseudo Match	age
checking	Copy	-	checking
savings	Copy	-	savings
income	Copy	mSeg Lookup	income
total_debt	Copy	mSeg Lookiup	total_debt
loan_payment_score	Copy	mSeg Lookup	loan_payment_score

These pseudonyms used to find the matching mSeg values for income, total_debt and loan_payment_score

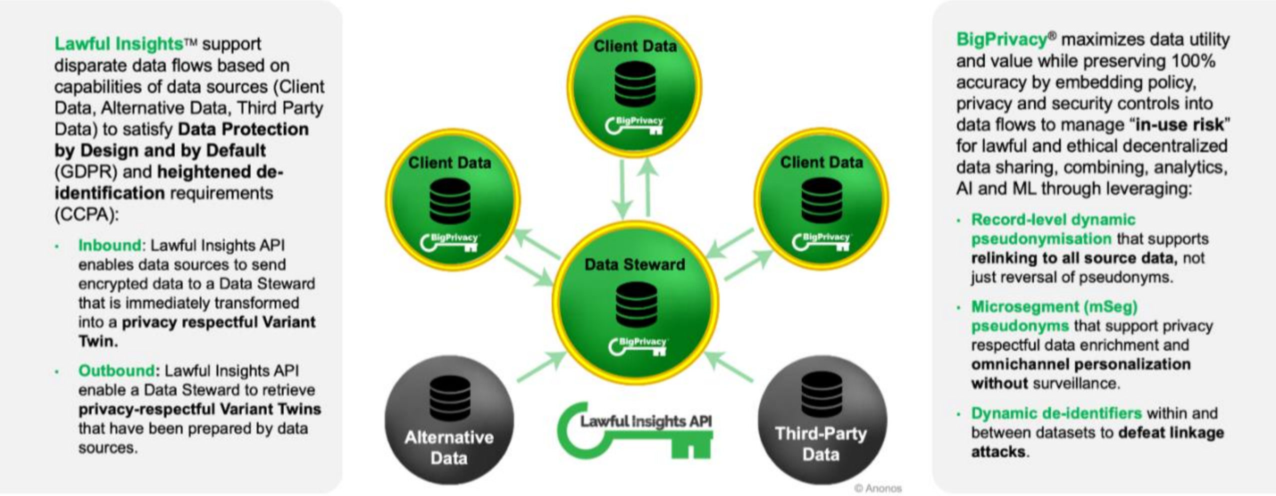
A R-DDID	B account_id	C name
RD-d286aceeb1023642a7cde911b3df10c6	account-CYEUuWQ10	Acosta

D gender	E age	F checking	G savings	I income	J total_debt	K loan_payment_score
M	72	1980	59597	70323	75312	121

mSeg defining fields in a source data set are transformed to deterministic pseudonyms (A-DDID's) that correspond to identical A-DDID's in an mSeg Variant Twin that has been prepared for sharing so that it can be used for enrichment. Pseudonym-to-pseudonym matching is then used to enrich the source data set with attribute data in the mSeg Variant Twin.

LAWFUL INSIGHTS API

Lawful Insights API is a special purpose application of Anonos technology designed for use in protecting data at the edge of a network – outbound by a sending party, or inbound by a receiving party. **Utilising the same processes, techniques and technology as BigPrivacy, it leverages API endpoints to reduce friction and to streamline and accelerate the process of safely and securely sending and receiving data for sharing, combining and enrichment, delivering Speed To Insight, Lawfully and Ethically.**



Lawful Insights™ support disparate data flows based on capabilities of data sources (Client Data, Alternative Data, Third Party Data) to satisfy **Data Protection by Design and by Default** (GDPR) and **heightened de-identification** requirements (CCPA):

- Inbound: Lawful Insights API enables data sources to send encrypted data to a Data Steward that is immediately transformed into a **privacy respectful Variant Twin.**
- Outbound**:** Lawful Insights API enable a Data Steward to retrieve **privacy-respectful Variant Twins** that have been prepared by data sources.

BigPrivacy® maximizes data utility and value while preserving 100% accuracy by embedding policy, privacy and security controls into data flows to manage **"in-use risk"** for lawful and ethical decentralized data sharing, combining, analytics, AI and ML through leveraging:

- Record-level dynamic pseudonymisation that supports **relinking to all source data,** not just reversal of pseudonyms.
- Microsegment (mSeg) pseudonyms that support privacy respectful data enrichment and **omnichannel personalization without** surveillance.
- Dynamic de-identifiers within and between datasets to **defeat linkage attacks.**

When Lawful Insights API is used by a sending party in an "outbound" mode, they first create a Variant Twin for sharing as described above.

The receiving party is then authorised using their instance of Big Privacy to access an API endpoint at the sender's instance that enables them to retrieve that specific Variant Twin. That Variant Twin is transmitted to them using standard https:// using TLS encryption, and is imported into their instance of BigPrivacy as what is known as a Shared Data set.

When Lawful Insights API is used in "inbound" mode, a receiving party first exposes an API endpoint in their instance of Lawful Insights API to a sending party allowing them to transmit via https:// a schema that describes data to be sent. The receiving party then configures a Privacy Transformer that will adequately protect to their requirements.

In this case, no master index will be created to ensure the receiving party can show that they are not in a position to reidentify data subjects.

The receiving party then notifies the sending party to use a second API endpoint to transmit data via https://. Upon receipt in memory the data is immediately transformed into Variant Twin form and stored as desired.

See the discussion of "Data Safe Haven #5: Expanded Data Use, Sharing & Combining" in the COMPLIANCE section below for more information about the significant benefits, including reductions in obligations and liabilities, for parties receiving data when using Anonos Lawful Insights API.

COMPLIANCE

Centralised approaches to data protection create tensions between the obligation of compliance teams to protect their organisation against threats, liability and business disruptions from data misuse, and the desire of business teams to generate digital insights. Many compliance teams advise the use of anonymisation to reconcile these tensions. **However, as explained in detail in the TECHNOLOGY section, anonymisation:**

- **Only works for centralised processing; and**
- **Results in lost linkability and the full context of data necessary for sophisticated data analytics, AI, ML, sharing, combining or enriching.**

Business teams are looking for digital insights that are sometimes only available from decentralised processing. This results in increasingly widespread practices of sharing, combining and enriching data with customer, partner or third-party data sources.

Compliance teams cannot claim the benefits of anonymisation in good faith when decentralised processing makes it impossible to be aware of "…all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly" as required for data to satisfy the requirements for anonymisation under the GDPR.[34]

Anonos has spent the past eight years exploring, in depth, the GDPR and predecessor EU data protection laws to understand "Data Safe Havens" – the following explicitly recognised combinations of GDPR legal & technical safeguards that maximise **Speed To Insight, Lawfully & Ethically.**

These Data Safe Havens leverage GDPR-compliant principles of Pseudonymisation[35] and Data Protection by Design and by Default,[36] together with patented Anonos state-of-the-art decentralised

[34] See GDPR Recital 26.

[35] The term Pseudonymisation is used fifteen (15) times in the GDPR, compared Anonymisation which is used only three (3) times, and Encryption which is used only four (4) times in the GDPR. No other Privacy Enhancing Techniques (PETs) are referenced in the GDPR. Benefits of GDPR compliant Pseudonymisation include, but are not limited to, the following: (i) tipping the balance in favour of Legitimate Interests processing (Articles 5(1)(a), 6(1)(f), and WP29 WP 217); (ii) more flexible change of purpose (Article 5(1)(b), WP29 WP 203); (iii) more expansive data minimisation (Articles 5(1)(c), 89(1)); (iv) more flexible storage limitation (Articles 5(1)(e), 89(1)); (v) enhanced security (Articles 5(1)(f), 32); (vi) more expansive further processing (Article 6(4), WP29 WP 217); (vii) more flexible profiling (WP29 WP 251 rev.01 - Annex 1, Recital 71, Article 22); and (viii) ability to lawfully and ethically share, combine and enhance data (recitals 42 and 43, Articles 11(2), 12(2), WP29 WP259 rev.01).Further details about GDPR requirements for and benefits of Pseudonymisation are included in the TECHNOLOGY section above. Additional information concerning Pseudonymisation are available at www.Pseudonymisation.com. See also Appendix B for Cross Reference to ENISA Pseudonymisation Guidance.

[36] See GDPR Article 25.

data protection techniques that enable Controlled Linkable Data.[37] This combination does not require the loss of linkability and full context of data for sophisticated data analytics, AI, ML, sharing, combining, or enriching.

Anonos' patented BigPrivacy software and API-based solutions embed policy, privacy and security risk-based controls into data flows to support the GDPR Data Safe Havens described below.

Anonos allows organisations to comply with the GDPR while achieving Speed To Insight, Lawfully & Ethically. This is what differentiates Anonos and delivers substantial value to organisations capitalising on this advantage.

DATA SAFE HAVEN #1: How to Lawfully Process Pre-GDPR Data

Many organisations have historically relied on general broad-based consent as their lawful basis for collecting, storing and other processing of pre-GDPR EU personal data ("Legacy Data"). **However, under the GDPR it is no longer legal to possess, store (in *either* encrypted or unencrypted format) or process Legacy Data since such broad-based consent often does not satisfy the GDPR's (or local case law related) consent requirements.** The GDPR has no "grandfather provision" or "exemption" that allows for ongoing possession, storage or use of (now) unlawful Legacy Data.[38]

This exposes organisations to:

1. Injunctions ordering the immediate suspension of data processing;
2. Substantial fines for a failure to delete now illegal Legacy Data; and
3. The unwillingness of customers, partners and third parties to use, share or combine Legacy Data due to potential liability and disruption to operations.

The GDPR-audited Pseudonymisation capabilities of BigPrivacy SaveYourData® software enables data controllers to transform Legacy Data.[39] This transformation, together with appropriate Data Protection Impact Assessment processes[40] can enable data controllers to exercise their "one off" opportunity to transform Legacy Data into a state that supports Legitimate Interests processing, rather than requiring the wholesale deletion of decades worth of valuable information to the detriment of data controllers and society as a whole.

This means that data controllers can avoid:

1. Having to delete valuable Legacy Data;

[37] The concept of Controlled Linkable Data was presented at an International Association of Privacy Professionals (IAPP) program entitled General Data Protection Regulation (GDPR) Big Data Analytics featuring Gwendal Le Grand, Director of Technology and Innovation at the French Data Protection Authority—the CNIL, Mike Hintze, Partner at Hintze Law and former Chief Privacy Counsel and Assistant General Counsel at Microsoft, and Gary LaFever, CEO and General Counsel at Anonos and former law partner at Hogan Lovells (see https://www.anonos.com/iapp-gdpr-data-analytics-webinar-replay) and explained in a Whitepaper co-authored by Messrs. Hintze and LaFever entitled Meeting Upcoming GDPR Requirements While Maximizing the Full Value of Data Analytics (see https://papers.ssrn.com/sol3/papers. cfm?abstract_id=2927540)

[38] See GDPR Recital 171, Articles 94(1) and 99, and WP29 Guidelines on Consent WP259 rev.01 at pg. 30, footnote 76.

[39] See the announcement of the EuroPrivacy audit of SaveYourData software at https://www.prnewswire.com/news-releases/anonos-saveyourdata-software-officially-certified-by-europrivacy-meets-the-requirements-of-the-eu-general-data-protection-regulation-gdpr-300741945.html. Copies of the EuroPrivacy audit are available to select parties upon written request to Anonos.

[40] See GDPR Article 35.

2. The risk of injunctions ordering immediate suspension of data processing;
3. Exposure to significant fines; and
4. Lost value from not being able to use, share, combine or enrich Legacy Data with customers, partners or third parties.

The SaveYourData® capability of Anonos BigPrivacy software provides a means to lawfully and ethically save Legacy Data, while an organisation implements solutions to address processing issues to comply with GDPR requirements.

The dilemma for data controllers is how to retain valuable Legacy Data when it plays a crucial role in the controller's digital transformation program and data-centric projects like sophisticated data analysis, AI, ML, sharing, combining or enriching.

Under the GDPR, a controller may be able to transform Legacy Data and use a new legal basis of Legitimate Interests using SaveYourData GDPR-audited Pseudonymisation capabilities as the first step in legally continuing with its data-driven journey. ***This might be of particular interest in cases where the controller has not been successful in reconsolidating its consent mechanisms, particularly when Legitimate Interests may be a more appropriate legal basis.***

Further action will be necessary to make use of the data in compliance with Data Safe Havens to maximise the full value of Legacy Data, but the data controller may have more flexibility to arrange suitable processing and not be forced to delete valuable data.

See IDC Report: ***Anonos' SaveYourData – "Deep Freezes" Enterprises' Existing Personal Data Sets as They Plan Analytics Strategies***[41]

USE CASE: Lawful Processing of Pre-GDPR Clinical Trial Data

EU-based clinical trial studies conducted prior to May 25, 2018[42] relied on general broad-based informed consent to comply with then-current EU data protection laws. Although this consent might comply with informed consent requirements under member-state national clinical trial regulations, it might not in all cases satisfy the separate and new requirements of GDPR-compliant consent for data protection purposes.

If Legacy Data obtained during these pre-GDPR clinical trials is stored to adhere to (amongst others) clinical trial obligations, such storage may represent unlawful processing for GDPR purposes. However, if in addition to having secured informed consent for clinical trial purposes, Legacy Data is transformed to support the new legal basis of Legitimate Interests (cumulated with the basis for

[41] The IDC report is available at https://www.anonos.com/hubfs/IDC_Report_Anonos_SaveYourData.pdf.

[42] The EU adopted the General Data Protection Regulation (GDPR) in 2016 to replace the1995 Data Protection Directive. Parties were provided two years advance notice to ensure that they complied with new GDPR requirements starting on May 25, 2018 onward.

processing health data for statistical, historical and scientific research under Article 9 (2)(j)), then ongoing storage may be lawful under the GDPR.

More importantly, the GDPR has introduced a new scheme to the standards of data processing by allowing data to be used for secondary processing purposes. Data controllers may also investigate using Legacy Data for secondary processing leading to the creation of new scientific breakthroughs and discoveries.

Rather than losing access to valuable data due to obligations to delete Legacy Data, this data may be transformed to support Legitimate Interests processing by implementing technical and organisational safeguards that meet the required legal and ethical requirements under the GDPR.

DATA SAFE HAVEN #2: Legitimate Interests Lawful Basis

There are significant questions as to the legality of consent as a valid basis under the GDPR for sophisticated data analysis, AI, ML, sharing, combining, or enriching.[43] This affects the value of these projects and the digital insights that can be extracted from them lawfully and ethically. **Anonos BigPrivacy leverages regulatory requirements as a competitive advantage to balance the increasing demand for digital insights.** In many situations, organisations can overcome the limitations of consent by using GDPR-compliant Pseudonymisation to enable Legitimate Interests as a lawful basis to support processing:

1. That cannot be described with required specificity at the time of initial data collection.
2. To avoid having to request re-consent each time a different processing of data is desired.
3. To avoid disruption to processing triggered by the revocation of consent.

For sophisticated data analysis, AI, ML, sharing, combining, or enriching to be legal under the GDPR (and evolving “GDPR-like” data protection laws), technology and organisational safeguards are required that support the requirements for Legitimate Interests processing. These safeguards must support the “Balancing of Interests” test and other tests necessary for valid Legitimate Interests-based processing.

Anonos BigPrivacy supports dynamism, functional separation[44] and other requirements for GDPR compliant Legitimate Interests processing, putting in place the necessary technical safeguards, and has received nine granted international patents in recognition of the utility and novelty of its inventions in this area.[45]

A critical component of the patented BigPrivacy process is supporting the use of the Legitimate Interests lawful basis, so that the creation, use, sharing, combining and enriching of BigPrivacy

[43] See the discussion regarding Limitations of Consent in the TECHNOLOGY section above.

[44] See the discussion regarding Functional Separation in the TECHNOLOGY section above.

[45] Anonos has been actively engaged in research and development to advance the state-of-the-art in data protection, privacy and security since 2012. Anonos’ decentralised data protection systems, methods and devices are covered by foundational granted patents (including, but not limited to: EU 3,063,691 issued in 2020; US 10,572,684 issued in 2020; CA 2,929,269 issued in 2019; US 10,043,035 issued in 2018; us 9,619,669 issued in 2017; US 9,361,481 issued in 2016; and US 9,129,133; 9,087,216; and 9,087,215 issued in 2015) and a portfolio of over 70 pending domestic and international patent applications. See Appendix A for more information on the Anonos Patent Strategy and Portfolio.

Variant Twin data assets is GDPR-compliant (see the TECHNOLOGY section above for further details).

Lawful Legitimate Interests processing requires more than mere words and "cannot be equated to the interest of companies to make a profit from our personal data" as made clear in the case filed by Privacy International against Acxiom and Oracle (data brokers), Equifax and Experian (credit reference agencies), and Criteo, Quantcast and Tapad (ad-tech companies) with data protection authorities in France, Ireland, and the UK.[46]

The Privacy International case makes it clear that to serve as a valid lawful basis, Legitimate Interests processing must satisfy a three-part test.

The first two tests are relatively easy to satisfy, but the third test requires technical and organisational safeguards to tip the balance in favour of the data controller. The three tests are:

1. **Legitimate Interests** test;
2. **Necessity** test; and
3. **Balancing of Interests** test which requires the application of technical and organisational safeguards to balance the interests of the data controller (or third party) against the individual data subjects' rights and freedoms.

Without technical and organisational safeguards that satisfy the Balancing of Interests test, many data processing activities that were commonly practiced for decades are no longer lawful under the GDPR (and might, in fact, never have been lawful in pre-GDPR times since the Legitimate Interests test has often not been executed correctly).

"Consent" as defined under the GDPR requires specificity that is often impossible to satisfy for sophisticated data analysis, AI, ML, sharing, combining, or enriching. It is also not possible to secure legally binding consent for processing activities in the future that cannot be described at the time of data collection. This makes it impossible to rely on consent as a lawful basis for many kinds of sophisticated data analysis, AI, ML, sharing, combining, and enriching that are critical for developing digital insights.

BigPrivacy uniquely helps to support Legitimate Interests processing, as a complement to consent, in numerous ways, including:

1. Patented dynamic de-identification functionality that separates information value from identity to defeat unauthorised reidentification between data sets via the Mosaic Effect;[47] and
2. Patented Variant Twin data that can be sourced, curated, combined, shared and processed on-premises and in the cloud, in compliance with applicable laws.

Benefits of BigPrivacy-enabled Legitimate Interests processing under the GDPR include the following:

[46] https://privacyinternational.org/advocacy/2434/why-weve-filed-complaints-against-companies-most-people-have-never-heard-and-what

[47] See www.MosaicEffect.com

- **Right to Restrict Processing**: If a data controller uses GDPR-compliant Legitimate Interests processing, under GDPR Article 18(1)(d) they will not have an obligation to comply with claims to restrict processing if it is proven that the controller's interests prevail over the rights and interests of the data subject.
- **Right to Data Portability**: Under GDPR Article 20(1), data controllers using Legitimate Interests processing are not subject to the right of portability which applies to processing based on consent or contract.
- **Right to Object**: Data subjects do not have the right to object to processing under GDPR Article 21(1) if a data controller uses GDPR-compliant Legitimate Interests processing and can prove compelling interests that override the rights and interests of the individual. However, data subjects always have the right under Article 21(3) to not receive direct marketing outreach resulting from data processing.

When using BigPrivacy to support Legitimate Interests processing:

- The data controller should put data subjects on notice **at the time of initial data collection** that:
  - It is relying on Legitimate Interests as a lawful basis for processing (e.g. to perform statistical analysis to improve product and service offerings, to enhance user experience, etc.);
  - State-of-the-art GDPR-compliant Pseudonymisation and other safeguards are used to support Legitimate Interests processing and ensure that the data controller's interests are balanced with the data subjects' interests by limiting any undue impact on the data subject; and
  - Data subjects have the unconditional right to opt out of receiving any direct marketing enabled by the Legitimate Interests processing.[48]
- The data controller should also document the results of the three-part test for Legitimate Interests outlined above, as evidence of greater accountability within its Data Protection Impact Assessment process.

USE CASE: Avoiding Undesirable Consequences of Withdrawal of Consent

EU clinical trial regulations deal with the withdrawal of consent by study participants with less far-reaching impacts than the separate issue of withdrawal of consent for purposes of the GDPR. The latter can potentially lead to the termination of entire studies based on a request for withdrawal by a single study participant. **This highly undesirable result can be avoided by opting for the legal basis of Legitimate Interests[49] for GDPR compliance purposes only, while separately complying with informed consent requirements for purposes of complying with clinical trial regulations.**

The reason for this tension between informed consent requirements under EU clinical trial regulations and GDPR requirements for consent is the imbalance of power between the participant and the sponsor/investigator of a clinical trial. This imbalance of power means that consent frequently cannot be "freely given" within the requirements of the GDPR.

[48] See GDPR Article 21(3)

[49] As supported by the European Data Protection Board in its Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR) Adopted on 23 January 2019. See https://edpb.europa.eu/our-work-tools/our-documents/opinion-art-70/opinion-32019-concerning-questions-and-answers-interplay_en

The European Data Protection Board (EDPB) considers that "*this will be the case when a participant is not in good health conditions, when participants belong to an economically or socially disadvantaged group or in any situation of institutional or hierarchical dependency.*"[50]

Rather than relying on GDPR-compliant consent in situations where the national requirements for clinical trial informed consent are too broad or vague because they were not designed under GDPR principles, the EDPB recommends the legal basis of Legitimate Interests for GDPR purposes.[51] So long as national requirements for clinical trial informed consent are satisfied, Legitimate Interests may be used as a legal basis for GDPR data protection purposes which, combined with Article 9(2)(j) provided there is a national law in place,[52] helps to avoid the undesirable consequences of the revocation of GDPR-based consent. However, as described above, one can only use Legitimate Interests for GDPR purposes if appropriate technical and organisational safeguards are implemented.

Anonos BigPrivacy enables all parties involved in processing clinical trial data to ensure that under the GDPR the processing of data can be done under the lawful basis of Legitimate Interests by providing state-of-the-art technical safeguards.

DATA SAFE HAVEN #3: Lawful Secondary Processing

Anonos BigPrivacy enables organisations to ensure secondary processing is "compatible" with the original primary purpose through Pseudonymisation and functional separation[53] of personal data. Under the GDPR, when an organisation processes personal data obtained for a particular permitted purpose, it cannot process it further except for purposes that are compatible.[54]

However, "further processing" of personal data may be deemed compatible with the original purpose if the processing satisfies the requirements of:

- **Article 6(4)** with respect to further "processing for a purpose other than that for which the personal data have been collected…not based on the data subject's consent"; or

- **Article 89(1)** with respect to processing conducted for "archiving purposes in the public interest," "scientific or historical research purposes," or "statistical purposes." The GDPR highlights Pseudonymisation as a safeguard to help ensure that such further processing is lawful.[55]

[50] Id at p.6.

[51] "*The EDPB considers that as an alternative to data subject's consent, the lawful grounds of processing provided under Article 6(1)(e) or 6(1)(f) are more appropriate. [...] For all other situations where the conduct of clinical trials cannot be considered as necessary for the performance of the public interest tasks vested in the controller by law, the EDPB will consider that*

the processing of personal data could be "necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject" following Article 6(1)(f) GDPR" (Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR) Adopted on 23 January 2019, p.7).

[52] Article 9(2)(j) GDPR provides for the cumulative legal basis necessary to process special categories of data (such as medical data) for statistical, historical and scientific research.

[53] See discussion regarding Functional Separation in the TECHNOLOGY section above.

[54] See GDPR Article 5(1)(b)

[55] See GDPR Articles 6(4)(e) and 89(1)

Further processing for “archiving purposes in the public interest,” “scientific or historical research purposes,” or “statistical purposes” is specifically considered not to be incompatible with the initial purposes if appropriate safeguards for data subjects are provided to ensure, in particular, data minimisation. Measures may include Pseudonymisation.[56] Pseudonymisation is also an explicitly-recognised safeguard under Article 6(4)(e) to help ensure that any such further processing of personal data “[are] compatible with the purpose for which the personal data are initially collected” in compliance with Article 5(1)(b) (“purpose limitation”) requirements. Accordingly, Anonos BigPrivacy’s Pseudonymisation and functional separation capabilities help to enable lawful and ethical secondary processing.

If organisations cannot process data for compatible secondary processing and “statistical purposes” to perform predictive analytics, then huge potential benefits are lost for data subjects, data controllers and society as a whole. Southampton University (UK) professors highlight the benefits of “a more constructive interpretation of the GDPR…on the basis of a dynamic approach to data protection law” that distinguishes between three different “…compliance stages (data collection, data analytics, individual impact)….”[57] Adopting this three-stage perspective with the capabilities of Anonos BigPrivacy in mind:

1. **Data Collection Stage:** Anonos BigPrivacy supports Legitimate Interests-based data collection;

2. **Data Analytics Stage**: BigPrivacy supports creation of non-identifying, dynamically de-identified derivative versions of original data referred to as “Variant Twins” for analysis; and

3. **Individual Impact Stage**: After the data controller has gathered, normalised, and analysed non-identifying Variant Twin data “in a way that equally respects their marketing interests and the privacy of users at large,”[58] then the Pseudonymisation/de-identification rules can be reversed under privacy-respectful controlled conditions to enable outreach to data subjects based on Legitimate Interests or consent.

USE CASE: Banking: Lawful Marketing (Secondary Processing) to Customers

In this example use case, a Bank collects personal data from customers and puts the customers on notice at the time of initial data collection that the Bank relies on Article 6(1)(f) Legitimate Interests processing to perform statistical analysis to improve product and service offerings for customers and to enhance future user experience.

This is done by leveraging GDPR-certified Pseudonymisation to ensure that the data controller’s interests are balanced with the data subjects’ rights and interests and by enforcing safeguards that limit the undue impact on the data subjects.

The Bank desires to use the data to improve product and service offerings and to enhance the user experience, including data that was previously acquired from ex-customers of the Bank. The Bank undertakes the following analysis under GDPR Article 6(4) to evaluate the lawfulness of using the data

[56] See GDPR Article 89(1)

[57] See *“Data Analytics and the GDPR: Friends or Foes? A Call for a Dynamic Approach to Data Protection Law”* by Sophie Stalla-Bourdillon and Alison Knight at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3248976 at page 16.

[58] See https://www.exchangewire.com/blog/2017/12/04/probabilistic-key-unlocks-new-markets/

collected to ensure that use of such personal data is compatible for the purpose for which the data were initially collected.

Article 6(4)(a)-(e) Further Processing Analysis:

a. *any link between the purposes for which the personal data have been collected and the purposes of the intended further processing;*

   There is a direct link between the original use of the data from active customers and the intended further use of the data from past customers to improve Bank product and service offerings and to enhance the user experience for Bank customers overall.

b. *the context in which the personal data have been collected, in particular regarding the relationship between data subjects and the controller;*

   The data was initially collected in the context of a Bank-customer relationship which is compatible with the intended further use of data from past customers to improve Bank product and service offerings and to enhance the user experience for Bank customers overall.

c. *the nature of the personal data, in particular whether special categories of personal data are processed, pursuant to Article 9, or whether personal data related to criminal convictions and offences are processed, pursuant to Article 10;*

   The nature of the personal data in question does not fit within any of the special categories of data.

d. *the possible consequences of the intended further processing for data subjects;*

   The intended further use of the data from past customers to improve Bank product and service offerings and enhance the user experience for Bank customers overall will not have adverse impacts on the data subjects who were prior customers of the bank.

e. *the existence of appropriate safeguards, which may include encryption or pseudonymisation.*

   The Bank leverages BigPrivacy to ensure that the data controller’s interests are balanced with the data subjects’ rights and interests by enforcing safeguards to limit undue impact on the data subjects by supporting GDPR-compliant principles of Pseudonymisation and Data Protection by Design and by Default, together with patented Anonos state-of-the-art decentralised data protection techniques.

Assuming the Bank properly documents the forgoing analysis under Article 6(4), it is reasonable to conclude that the desired further processing will be compatible with the purpose for which the personal data were initially collected.

DATA SAFE HAVEN #4: Data Protection by Design and by Default

The GDPR further imposes new requirements for Data Protection by Design and by Default which means organisations must integrate or 'bake in' significant data protection capabilities into processing practices, from the design stage through the full data lifecycle. **Previously known as 'Privacy by Design', this concept has long been part of data protection law. However, two key changes which are newly mandated under the GDPR are:**

1. It is now a legal mandate to support more than just privacy by design: Data Protection by Design and by Default requires **the most stringent implementation of privacy by design**; and

2. It has heightened requirements, including the need to support the GDPR principles of data minimisation and purpose limitation **to limit data use to the minimum extent and time necessary to support each specific product or service authorised by a data subject.**

The obligation to support Data Protection by Design and by Default as newly-defined under the GDPR requires each organisation "to be clear in advance about what its plans for secondary processing of personal data intends to achieve…[including] the upfront design of data processing to demonstrate that this thinking has taken place and to ensure safeguards measures can be implemented to mitigate any notable risk areas identified." [59] It also notes that "data minimization should be engineered relative to purposes before the start of processing, at the time of the determination of the means."[60]

This essentially means that less, rather than more, personal data must be provided, used or disclosed or otherwise processed for a given purpose. How much less? Only the **minimum amount** needed to achieve the authorised purpose. BigPrivacy supports GDPR-compliant Pseudonymisation, which dynamically enforces data minimisation via fine-grained access controls leveraging Controlled Linkable Data®.[61] This enables the disclosure of only the "minimum identifying data" to those who need to know, all on a case-by-case basis.

While the focus of data minimisation has usually been on minimising the amount of personal data **collected** at the acquisition stage, data minimisation also applies to the post-collection **use** of personal data. Accordingly, BigPrivacy helps support data use minimisation within an organisation by enforcing selective access to data, ensuring that an individual employee only has access to the data required for them to do their job and no more.

When personal data is shared between organisations, BigPrivacy enforces selective "in-use" risk management controls to ensure that data is used only as authorised.

[59] See GDPR Article 25

[60] See "*Data Analytics and the GDPR: Friends or Foes? A Call for a Dynamic Approach to Data Protection Law"* by Sophie Stalla-Bourdillon and Alison Knight at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3248976 at page 15.

[61] The concept of Controlled Linkable Data was presented at an International Association of Privacy Professionals (IAPP) program entitled General Data Protection Regulation (GDPR) Big Data Analytics featuring Gwendal Le Grand, Director of Technology and Innovation at the French Data Protection Authority—the CNIL, Mike Hintze, Partner at Hintze Law and former Chief Privacy Counsel and Assistant General Counsel at Microsoft, and Gary LaFever, CEO and General Counsel at Anonos and former law partner at Hogan Lovells (see https://www.anonos.com/iapp-gdpr-data-analytics-webinar-replay) and explained in a Whitepaper co-authored by Messrs. Hintze and LaFever entitled Meeting Upcoming GDPR Requirements While Maximizing the Full Value of Data Analytics (see https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2927540)

In summary, BigPrivacy helps to ensure that only discrete data elements are made available to support minimal authorised use.

BigPrivacy enforces "in-use" risk management by leveraging the most current information about a user, data and environment at the time of disclosure to dynamically enforce the appropriate level of data resolution, as if viewing the data through a "lens." The lower the magnification, the less identifying the disclosed data is (while still conveying necessary information value), whereas with higher magnification, the more "identifying" the disclosed data becomes. By leveraging BigPrivacy, only the minimum required level of identifying data is revealed for each authorised purpose. This unique resolution level of data disclosed via such "lens" is a Variant Twin of the original underlying data.

With BigPrivacy, productive data use can continue by disclosing Variant Twins[62] to convey only the necessary information value to accomplish permitted data processing in a privacy-respectful and non-identifying manner.

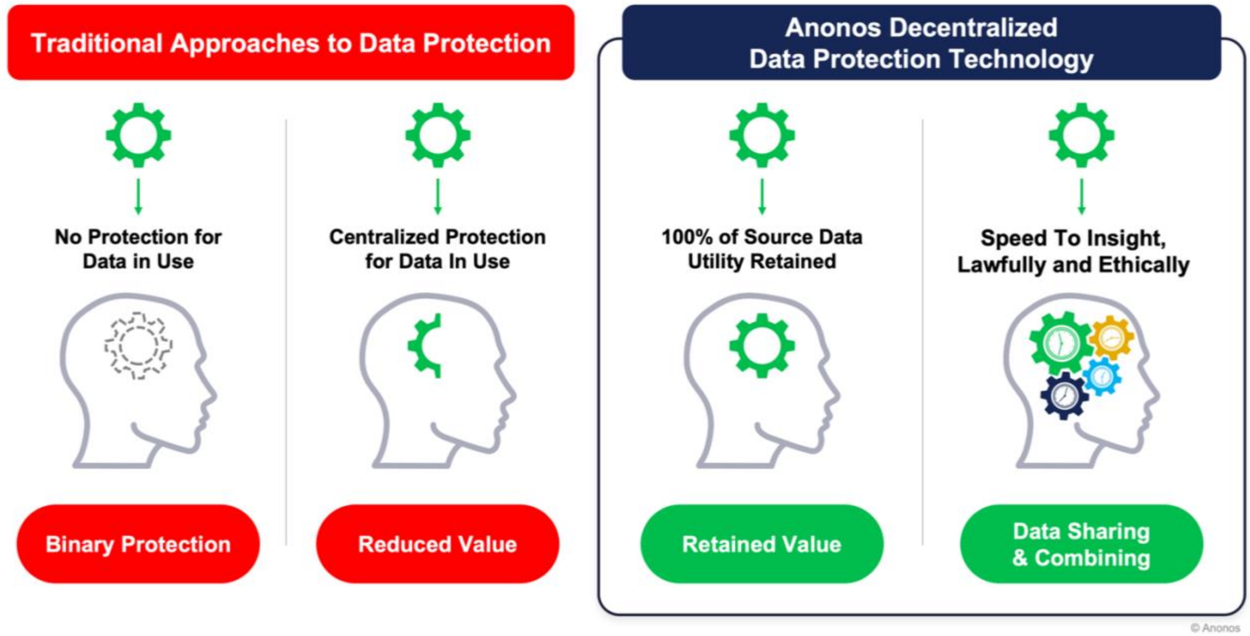
Over the full lifecycle of data, BigPrivacy-enabled Data Protection by Design and by Default:

- Maximises authorised uses of data while minimising unauthorised uses of data – all by minimising reidentification risks;
- Facilitates compliance with and auditability against privacy data protection policies by enabling the mathematical, statistical and/or actuarial measurement and monitoring of data use;
- Enables common data store(s) to simultaneously programmatically support data protection policies applicable to different use cases – and to do so simultaneously; and
- Adjusts in real-time to the changing requirements of policies by dynamically modifying the intelligible form of data into which protected data are transformed.

In the figure below, the first column represents the effect of binary security techniques (like encryption) where the top green gear reflects the value of original data (in unprotected form) and the empty gear below represents the value of the data when it is in a protected form, rendering it unusable since the data is unusable in its protected form.

The second column illustrates the reduction in data value – from the full green gear at the top to the half gear at the bottom representing reduced value because of: (i) restrictions of centralised-only data applications; (ii) the value-reducing impact of traditional approaches to data protection involving suppression, perturbation, masking and generalisation of data; and (iii) the removal of data from the ecosystem due to concerns over the lawful and ethical possession and use of data.

[62] See discussion regarding Variant Twins in TECHNOLOGY section above.


Traditional Approaches to Data Protection
Anonos Decentralized Data Protection Technology
No Protection for Data in Use
Centralized Protection for Data In Use
100% of Source Data Utility Retained
Speed To Insight, Lawfully and Ethically
Binary Protection
Reduced Value
Retained Value
Data Sharing & Combining
© Anonos

The third column shows how BigPrivacy Data Protection by Design and by Default capabilities enable retention of 100% original source data value represented by the full green gear.

The fourth column illustrates how BigPrivacy enables protection to dynamically flow with data to technologically enforce polices to enable “Speed To Insight, Lawfully and Ethically” represented by the interconnected gears of different colors.

USE CASE: Medical Research: Dynamic Data Use Controls

BigPrivacy-enabled Data Protection by Design and by Default empowers data scientists to move beyond the “Middle Ages” approach of repeated, bespoke manual assessments – analogous to monks’ manually copying manuscripts – to better leverage and scale data scientists’ expertise.
US National Institutes of Health (“NIH”) Director Elias Zerhouni once testified before Congress that Industrial Age medicine had focused on mass production of “one-size-fits-all” remedies often applied too late in the disease process but suggested that Information Age healthcare technologies could be predictive, preemptive, precise, and participative.[63]

To support Information Age healthcare, BigPrivacy-enabled Data Protection by Design and by Default can be embedded into data to reduce the risk of reidentification from sharing, combining and enriching.

BigPrivacy-enabled Data Protection by Design and by Default technology makes it much easier to establish that:

- The risk is very small that information which is technologically enforced via pre-approved policies could be used, alone or in combination with other reasonably available information, to identify an

[63] See http://www.nature.com/nbt/journal/v34/n5/full/nbt.3514.html

individual who is the subject of the information; and

- The methods and results underlying the analysis embodied in each policy are well documented.

These capabilities could provide greater flexibility to Institutional Review Boards (IRBs), Independent Ethics Committees (IECs) and Ethical Review Boards (ERBs) to help overcome the limitations of traditional approaches to medical research. As one example, a study of nearly 600,000 people found 13 surviving adults with genetic abnormalities from which most people die as children.[64] Each of these 13 individuals, therefore, represented an informational goldmine for developing breakthrough treatments (including orphan drugs) or cures to treat those afflicted with genetic abnormalities. But because of the binary, one-time consent the individuals provided, researchers were unable to identify any of the 13 people.

The use of technical and organisational safeguards enabled by BigPrivacy to enforce Data Protection by Design and by Default principles could have allowed the IRB (IEC or ERB) in this study to enable participants to authorise only specific uses of their data initially, with the flexibility to approve later further use in the service of developing breakthrough treatments or cures – while protecting the fundamental rights of the parties involved.

USE CASE: Self-Service Business Intelligence: Lawfully & Ethically[65]

Organizations depend on digital insights to achieve competitive advantage and industry leadership. An industry Whitepaper highlights the benefits of self-service business intelligence or “BI” in helping to meet the demand for timely data-driven insights:

> *“Instead of confining your organization to a small number of expensive, elite BI professionals, self-service BI equips individuals throughout the organization to investigate their own data, including creating reports and dashboards as well as ad-hoc analysis. The good news? These people are experts and they know exactly the questions they need to ask and answer. The great news? They’ll have a solution that gets them answers quickly that they, in turn, can share with colleagues, fueling the reality of a data-driven organization.”*[66]

Self-service BI is becoming more accessible to satisfy real-time enterprise demands for data. Employees can use self-service BI and visualization tools to perform data analysis and reporting on their own without relying on IT departments. However, self-service BI systems expose organizations to potential liability and disruption of operations if the processing does not comply with legal and ethical obligations.

[64] See https://www.nature.com/articles/nbt.3514

[65] Anonos is building out this capability by leveraging BigPrivacy technology and Intellectual Property (IP) rights with partners.

[66] See Tableau Software paper titled “*7 Signs You Need Self-Service Reports: How IT Can Empower Users—And Help Themselves In The Process*” at https://www.tableau.com/sites/default/files/media/whitepaper_7_signs_you_need_self-service_reports_0_0_1.pdf

Anonos patents[67] cover the use of in-use data protection risk-based controls to enforce Data Protection by Design and by Default principles to help ensure that self-service BI remains lawful and ethical.

Example principles include:

- Intercepting queries in real-time.
- Matching a query with policy, security and privacy risk-based controls (which may include any privacy enhancement techniques (PET), including data protection, dynamic de-identification, anonymity, pseudonymity, granularization, or obscurity policies).
- Configuring a Variant Twin to reveal only the data that is necessary for an authorized purpose, period, place or other criteria by obfuscating data values.
- Transforming the source data to comply with the Variant Twin requirements.
- Delivering the transformed Variant Twin to the user for self-service analytics in real-time.
- Implementing these capabilities in classical and quantum computing environments to overcome the increasing vulnerability of cryptographic algorithms, otherwise thought to be secure, that can be efficiently broken by a sufficiently powerful quantum computer.

DATA SAFE HAVEN #5: Expanded Data Use, Sharing & Combining

The GDPR clarifies and enhances the privacy rights of individual data subjects with new well-known rights such as the “right to be forgotten,” the “right to data portability” and more. However, under GDPR Articles 11(2) and 12(2), if the purposes for which an organisation processes personal data do not or no longer require identification of an individual, and the organisation can show **that it is not in a position to identify the data subject**, then it does not need to comply with these data subject rights.[68]

If personal data is pseudonymised using BigPrivacy, so that a given controller or processor cannot identify the individuals concerned, such organisation may not be subject to certain obligations. **BigPrivacy helps to limit the risk of using personal data by data controllers and processors down the data chain.** This enables the data to be used going forward in a “risk-reduced” manner, which dramatically limits the likelihood of data subjects being re-identified.

In addition, non-identifying Variant Twin versions of data processed under the lawful basis of Legitimate Interests become the proprietary data assets of an organisation, with respect to which there is no obligation to provide copies to third parties (which may be competitors – e.g., FinTechs under the second Payment Services Directive (PSD2), a European directive designed to boost competition and the variety of financial services offerings). In addition, the right of data

[67] For example, see Patent No. 10,043,035 titled “Systems and Methods for Enhancing Data Protection by Anonosizing Structured and Unstructured Data and Incorporating Machine Learning and Artificial Intelligence in Classical and Quantum Computing Environments.” See Appendix A for more information on the Anonos Patent Strategy and Portfolio.

[68] Subject to the right of a data subject under Article 11(2), for the purpose of exercising his or her rights, to provide additional information enabling his or her identification.

portability under GDPR Article 20 ***applies only to*** personal data processed using consent (under Article 6(1)(a) or Article 9(2)(a)) or contract (under Article 6(1)(b)), and not to data processed based on Legitimate Interests.[69]

As long as a controller can prove “compelling legitimate grounds for processing which override the interests, rights and freedoms” of data subjects due to the use of state-of-the-art technical and organisational safeguards (or “for the establishment, exercise or defence of legal claims”), objections by data subjects under Article 21 to using Variant Twin data for Legitimate Interests processing for sophisticated data analysis, AI, ML, sharing, combining, or enriching may be unsuccessful.[70]

In addition, GDPR-compliant Pseudonymisation can alleviate a data controller’s requirements to carry out data subjects’ rights of access under Article 15, rectification under Article 16, and erasure (“right to be forgotten”) under Article 17. Article 11 provides an exemption from these rights in the event that "the controller is able to demonstrate that it is not in a position to identify the data subject." Since the GDPR does not require a controller to hold additional information "for the sole purpose of complying with this Regulation," a data controller may use Pseudonymisation techniques and subsequently delete information that would enable the reversal of the pseudonymisation to identify individual data subjects.[71]

Data that is protected using BigPrivacy Lawful Insights API, as described in the TECHNOLOGY section above, may benefit from expanded use rights as outlined above.

USE CASE: Maximising Utility of Research Study Results

In the EU, hospitals and (academic) research centers are obliged when publishing the results of research studies in scientific journals to store the underlying data (often containing identifiable GDPR Article 9 “special category” health data) in a regulated data repository. This is to ***enable and allow*** other researchers to (i) verify the study results and (ii) use the data for further research activities. Confronted with this obligation, research institutions often ***anonymise the data***, reducing its utility and value to near-zero. In doing so, they also violate the obligation to enable others to verify study results.

Research institutions would benefit from GDPR-compliant Pseudonymisation using BigPrivacy when transferring such data sets to a regulated data repository. This would safeguard data utility and value for further expanded use, sharing, combining, and enriching. They could then comply with the obligation to minimise the risk that further use from data sharing, combining and enriching can entail (e.g. ensuring data minimisation, protecting the rights of data subjects, proving demonstrable accountability, etc.).

As explained above, BigPrivacy-enabled, GDPR-compliant Pseudonymisation dynamically enforces data minimisation via fine-grained risk-controls leveraging Controlled Linkable Data.[72] This enables

[69] See GDPR Article 20(1)

[70] See GDPR Article 21(1)

[71] Id

[72] The concept of Controlled Linkable Data was presented at an International Association of Privacy Professionals (IAPP) program entitled General Data Protection Regulation (GDPR) Big Data Analytics featuring Gwendal Le Grand, Director of Technology and Innovation at the French Data Protection Authority—the CNIL, Mike Hintze, Partner at Hintze Law and former Chief Privacy Counsel and Assistant General Counsel at Microsoft, and Gary LaFever, CEO and General Counsel at Anonos and former law partner at Hogan Lovells (see https://www.anonos.com/iapp-gdpr-data-analytics-webinar-replay) and explained in a Whitepaper co-authored by

the disclosure of only the "minimum identifying data" to those with a need to know, all on a case-by-case basis. By using BigPrivacy-enabled, GDPR-compliant Pseudonymisation, the data controller can protect data while it is in use without compromising the value and utility of the data for further extended use, sharing, combining and enriching purposes.

DATA SAFE HAVEN #6: Compliant Cloud Processing Under GDPR

Prior to the GDPR, only customers of Cloud Service Providers (CSPs) had direct liability for non-compliance under EU data protection laws. **That changed under the GDPR**, which introduced direct obligations, liability and exposure that **cannot** be negotiated away in contracts between CSPs and customers.

CSPs providing services involving EU personal data now have direct obligations, liability and exposure under the GDPR.[73] In addition, data controllers have an affirmative obligation to "use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of [the GDPR] and ensure the protection of the rights of the data subject."[74]

GDPR Recital 28 specifically highlights the benefits of GDPR-compliant Pseudonymisation, such as is enabled by BigPrivacy, including in the context of CSP "as-a-service" offerings. Recital 28 states that "the application of Pseudonymisation to personal data can reduce the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations."

Recital 78 goes even further by stating, "When developing, designing, selecting and using applications, services and products that are based on the processing of personal data or process personal data to fulfil their task, producers of the products, services and applications should be encouraged to take into account the right to data protection when developing and designing such products, services and applications and, with due regard to the state of the art, to make sure that controllers and processors are able to fulfil their data protection obligations."

The Cloud Security Alliance (CSA) Code of Conduct for GDPR Compliance highlights the importance of these issues in the requirement that:

> *"A pre-condition for relying on cloud computing arrangements is for the controller [cloud client] to perform an adequate risk assessment exercise, including the locations of the servers where the data are processed and the consideration of risks and benefits from a data protection perspective."*[75]

Messrs. Hintze and LaFever entitled Meeting Upcoming GDPR Requirements While Maximizing the Full Value of Data Analytics (see https://papers.ssrn.com/sol3/papers. cfm?abstract_id=2927540)

[73] See GDPR Article 82

[74] See GDPR Article 28(1)

[75] See https://www.anonos.com/cloud_security_alliance_code_of_conduct at page 2

The importance of these matters is further emphasised by the following assessment by the CSA of the decision by the Court of Justice of the European Union (CJEU) in the 2018 *Wirtschaftsakademie* case (CJEU Case C-210/16):

> *CSPs should examine carefully the relationship they have with their cloud customers, in order to accurately determine the role which each party plays regarding a given service. [The Wirtschaftsakademie] decision has vastly expanded the understanding of how "joint controllership" should be interpreted, and there may be cases where a CSP previously considered itself as acting as an autonomous controller (e.g., because it uses data provided by a cloud customer for a purpose defined by the CSP) which may, effectively, be more appropriately classified as a case of joint controllership (e.g., potentially, where the processing carried out by the CSP is actually done in order to improve the services provided to a customer).*[76]

Data controllers and CSPs bear (effective) “joint and several liability” to compensate data subjects for their material and non-material (non-monetary losses like damage to reputation, emotional distress, pain and suffering, etc.) damage, even if other parties in the supply chain were more at fault. This is because the aim of the GDPR is to ensure data subjects are made whole for any loss or damage they suffer.[77]

USE CASE: GDPR Compliant Cloud Processing[78]

Compliance with the GDPR for cloud processing is complicated. A Computerworld article reported that only 12% of 177 global IT organisations understood how GDPR affects cloud services.[79] The *Wirtschaftsakademie* ruling by the CJEU emphasises the importance of data controllers and CSPs being joint controllers. As joint controllers, neither the original data controller nor the CSP can abdicate its GDPR compliance obligations via contractual allocation of risk between the parties.

The *Wirtschaftsakademie* ruling by the CJEU stands for the proposition that joint responsibility does not require that each of the controllers have access to personal data processed, and yet they share responsibility for misuse.[80]

[76] See https://www.anonos.com/cloud_security_alliance_code_of_conduct at page 25

[77] See, inter alia, GDPR Recitals 13, 18, 22, 23, 24, 28, 36, 77, 78, 79, 80, 81, 82, 83, 95, 101, 108, 114, 115, 122, 124, 126, 127, 131, 145, 146, and Articles 3(1), 3(2), 4(8), 27, 28, 29, 30(2), 31, 32, 33, 35, 36, 37, 38, 39, 44, 46, 58, 79, 82, and 83. See also Hon, W. Kuan. “*Data Localization Laws and Policy: the EU Data Protection International Transfers Restriction through a Cloud Computing Lens.*” Edward Elgar Publishing, 2017.

[78] Anonos is building out this capability by leveraging BigPrivacy technology and Intellectual Property (IP) rights with partners.

[79] See https://www.computerworld.com/article/3427507/how-to-ensure-gdpr-compliance-in-the-cloud.html

[80] See https://www.twobirds.com/en/news/articles/2018/global/what-is-next-after-the-ecj-ruling-on-joint-control The *Wirtschaftsakademie* ruling was nuanced by the CJEU decision in *Fashion ID v Verbraucherzentrale* (CJEU Case C-40/17), where it was determined that each of joint controllers is only responsible for the activities where it effectively codetermines the means and purposes of the processing. Article 26(3) of the GDPR provides that “irrespective of the terms of the arrangement referred to in paragraph 1, the data subject may exercise his or her rights under this Regulation in respect of and against each of the controllers.” Article 82(3) limits the liability of a controller or processor “if it proves that it is not in any way responsible for the event giving rise to the damage” *(emphasis added).* So, if the CSP is a joint controller, rather than simply a processor following instructions from the customer with regard to the essential elements of the processing, then both the CSP and the customer may be jointly and

The *Wirtschaftsakademie* ruling highlights the importance of creating Anonos Variant Twins so that an original data controller retains control over the re-linkability of data to avoid potential liability. By using BigPrivacy, a data controller can share only Variant Twin data with third party “as-a-service” CSP providers, retaining control over identity re-linkability to reduce the risk of unauthorised reidentification. This benefits both the original data controller and the CSP as a result.

Anonos BigPrivacy enables more effective use of cloud computing ecosystems by protecting against liability exposure and risk of business interruptions due to failure to comply with complex cloud related GDPR compliance obligations.

severally liable. However, if the customer can show that the unlawful processing is entirely due to the CSP, then GDPR Article 82(3) would absolve the customer of liability.

KEY TAKEAWAYS

This Blueprint describes how Anonos BigPrivacy software, Lawful Insights API and Anonos Intellectual Property (IP) can embed policy, privacy and security risk-based controls into data flows to manage risk while data is in use.

Anonos solves one of the biggest challenges to successful digital transformation: balancing data utility and data protection to enable maximum value from sophisticated analytics, AI, machine learning (ML), data sharing, combining, and enriching in real-time while ensuring that data is protected in use.

Approaching data enablement technology solutions with an approach based on functional separation, Anonos has implemented a combination of anonymisation techniques, GDPR-compliant Pseudonymisation, k-anonymity, and patented Controlled Linkable Data to produce privacy-respectful Variant Twins. These Variant Twins enable the use, sharing, and combining of data assets in distributed and decentralised processing environments.

For an organisation to be successful, it must harmonise data protection and data use by fully involving and aligning different stakeholder groups within the organisation. To that end**, this Blueprint covers the perspective of Business, Technology and Compliance, to show how Anonos BigPrivacy is the state-of-the-art in technology enabling Speed To Insight, Lawfully and Ethically.**

APPENDIX A

ANONOS PATENT STRATEGY & PORTFOLIO

Anonos employs the following principles when developing and protecting Intellectual Property (IP) assets to ensure that customers and partners benefit from the uninterrupted ability to embed policy, privacy and security risk-based controls into data flows to protect both direct and indirect identifiers when data is used in decentralised processing – all while preserving 100% of source data value.

1. **"Biopharma Approach"** – From the very beginning, Anonos has pursued a "biopharma approach" to innovation seeking protection for as foundational and widespread IP as possible. In the early years, we always worked with two different law firms at the same time, alternating between one law firm helping to draft each patent specification, drawings and claims and a second law firm critiquing the work of the first firm and acting as if they represented a fictitious competitor looking to invalidate the patent. In addition, Anonos engaged expert parties whose practice is to provide evidence and testimony to disqualify patents. Anonos regularly engages experts to anticipate and overcome potential shortcomings that might later be used to invalidate granted patents.

2. **Provisional Patent Filings** – Provisional patent applications are filed on a regular basis to cover material improvements in innovation. Each year, we file at least one utility filing that incorporates by reference all of the provisional filings during the preceding one-year period so that we do not lose the protection of innovations disclosed in the provisional filings. In doing this, we have the ability to go back and mine from provisional filings for additional capabilities and coverage.

3. **In-Depth Specification and Drawings** – The most recently granted Patent No. 10,572,684 is 137 pages long, including an in-depth specification and drawings which have not yet been referenced in filed claims. Anonos will mine the specification and drawings for new claims covering additional capabilities with effectives dates that will go back to the date of the original filings.

4. **Keep Patent Family Alive to Enable Expanded Coverage** – Anonos always keeps at least one patent in active prosecution to "keep the family alive" to ensure that we can add additional coverage as per #2 and #3 above.

5. **International Protection** – Anonos has nine granted patents: seven in the US; one in the EU; and one in Canada. International protection for the subject matter of all of the granted US patents is in process. Anonos has 70+ additional patent filings.

Granted Patents

Systems and Methods for Enforcing Centralized Privacy Controls in De-Centralized Systems: US 10,572,684 (2020); international patents pending.

Systems, computer-readable media, and methods for improving both data privacy/anonymity and data value, wherein data related to a data subject can be used and stored, e.g., in a distributed ledger data structure, such as a blockchain, while minimizing reidentification risk by unauthorized parties and enabling data, including quasi-identifiers, related to the data subject to be disclosed to any authorized party by granting access only to the data relevant to that authorized party's purpose, time period, place and/or other criterion via the obfuscation of specific data values, e.g., pursuant to the European Union's General Data Protection Regulation (GDPR) or other similar regulatory schemes. The techniques described herein maintain this level of privacy/anonymity while still satisfying the immutability, auditability, and verification mandated by blockchain and other distributed ledger technologies (DLTs) for the decentralized storage of transactional data. Such systems, media, and methods may be implemented on both classical and quantum computing devices.

Dynamic De-Identification and Anonymity: EU 3,063,691 (2020); CANADA 2,929,269 (2019); US 9,129,133; 9,087,216; and 9,087,215 (2015).

Various systems, computer-readable media, and computer-implemented methods of providing improved data privacy, anonymity and security by enabling subjects to which data pertains to remain “dynamically anonymous,” i.e., anonymous for as long as is desired—and to the extent that is desired—are disclosed herein. Embodiments include systems that create, access, use, store and/or erase data with increased privacy, anonymity and security, thereby facilitating the availability of more qualified and accurate information. When data is authorized by subjects to be shared with third parties, embodiments may facilitate sharing information in a dynamically controlled manner that enables delivery of temporally-, geographically-, and/or purpose-limited information to the receiving party. In one example, anonymity measurement scores may be calculated for the shared data elements so that a level of consent/involvement required by the Data Subject before sharing the relevant data elements to third parties may be specified.

Systems and Methods for Enhancing Data Protection by Anonosizing Structured and Unstructured Data and Incorporating Machine Learning and Artificial Intelligence in Classical and Quantum Computing Environments: US 10,043,035 (2018); international patents pending.

Systems, computer-readable media, and methods for improving both data privacy/anonymity and data value, wherein real-world, synthetic, or other data related to a data subject can be used while minimizing reidentification risk by unauthorized parties and enabling data, including quasi-identifiers, related to the data subject to be disclosed to any authorized party by granting access only to the data relevant to that authorized party's purpose, time period, purpose, place and/or other criterion via the required obfuscation of specific data values, e.g., pursuant to the GDPR or HIPAA, by incorporating a given range of those values into a cohort, wherein only the defined cohort values are disclosed to the given authorized party. Privacy policies may include any privacy enhancement techniques (PET), including data protection, dynamic de-identification, anonymity, pseudonymity, granularization, and/or obscurity policies. Such systems, media and methods may be implemented on both classical and quantum computing devices.

Systems And Methods For Anonosizing Data: US 9,619,669 (2017); international patents pending.

Various systems, computer-readable media, and computer-implemented methods of providing improved data privacy, anonymity, and security by enabling subjects to which data pertains to remain “dynamically anonymous,” i.e., anonymous for as long as is desired—and to the extent desired—are disclosed herein. This concept is also referred to herein as “anonosizing.” In some embodiments, the anonosizing of data may be implemented by encoding and decoding data under controlled conditions to support specific uses within designated authorized contexts. By anonosizing data controls via “identifying” and/or “associating” data elements within a population, data uses may be restricted to only those uses permissioned by a data subject or authorized third party. If new authorized data uses arise, all original data value and utility may be retained to support them—to the extent authorized by a data subject or authorized third party—but inappropriate, i.e., non-permissioned, uses of identifying information may be prevented.

Systems and Methods for Contextualized Data Protection: US 9,361,481 (2016); international patents pending.

Various systems, computer-readable media, and computer-implemented methods of providing improved data privacy, anonymity, and security by enabling subjects to which data pertains to remain “dynamically anonymous,” i.e., anonymous for as long as is desired—and to the extent that is desired—are disclosed herein. This concept is also referred to herein as Just-In-Time-Identity, or “JITI.” Embodiments include systems that create, access, use, store and/or erase data with increased privacy, anonymity and security—thereby facilitating the availability of more qualified information—via the use of temporally unique, dynamically changing de-identifiers (“DDIDs”). In some embodiments, specialized JITI keys may be used to “unlock” different views of the same DDID (or its underlying value), thereby providing granular control over the level of detail or obfuscation visible to each user based on the context of said user's authorized use of data, e.g., authorized purpose(s), place(s), time(s), or other attributes of the use.

APPENDIX B

CROSS REFERENCE TO ENISA PSEUDONYMISATION GUIDANCE [81]

ENISA GUIDELINE NO 1 – NOVEMBER 2018: Recommendations on Shaping Technology According to GDPR Provisions – An Overview on Data Pseudonymisation[82]	Section	Anonos BigPrivacy
Personal Identifiers Replaced with Pseudonyms	2.1.1	✓
Pseudonyms Do Not Allow the Direct Derivation of Personal Identifiers	2.1.1	✓
Personal Data Can No Longer Be Attributed to a Specific Data Subject Without the Use of Additional Information	2.1.2	✓
Reversal of Pseudonymisation Is Non-Trivial in Absence of Additional Information	2.1.2	✓
Additional Information Kept Separately Using Technical and Organizational Controls to Limit Access	2.1.2	✓
Pseudonyms Applied to Direct and Indirect Identifiers	2.1.2, 2.1.3	✓
Resistance Against Re-Identification Via Singling Out	2.1.2, 2.1.3	✓
Resistance Against Re-Identification Via Linkage Attacks	2.1.2, 2.1.3	✓
Resistance Against Re-Identification Via Inference Attacks	2.1.3, 2.2	✓
Anonymisation Techniques Used to Further Reduce the Possibility of Third Parties Inferring Identity	2.2	✓
Single Input Results in a Decoupled Pair of Outputs: Pseudonymous Data and Additional Information Necessary to Reidentify	2.3	✓
Identify of Data Subjects Hidden in the Context of a Specific Data Processing Operation	2.3	✓
Any Recipient or Third-Party Having Access to Pseudonymised Data Cannot Trivially Derive Original Data Set and Identity of Data Subjects	2.3	✓
Support for Unlinkability Across Different Data Processing Domains	2.3	✓
Support for Accuracy by Retaining Access to Both Pseudonymised Output and Additional Information Necessary to Reidentify	2.3	✓
Does Not Use Hashing Without Key or Salt to Generate Pseudonyms	3.2	✓
Offers Keyed Hash Function (HMAC, SHA2/3, 256+ bit keys) to Generate Pseudonyms	3.3	✓
Uses Symmetric Encryption to Generate Pseudonyms	3.6	✓
Offers Tokens (Randomly Generated Values) As Pseudonyms	3.6	✓

[81] References to ENISA do not indicate any relationship, sponsorship, or endorsement by ENISA. All references to ENISA are intended to constitute nominative fair use under applicable trademark laws.

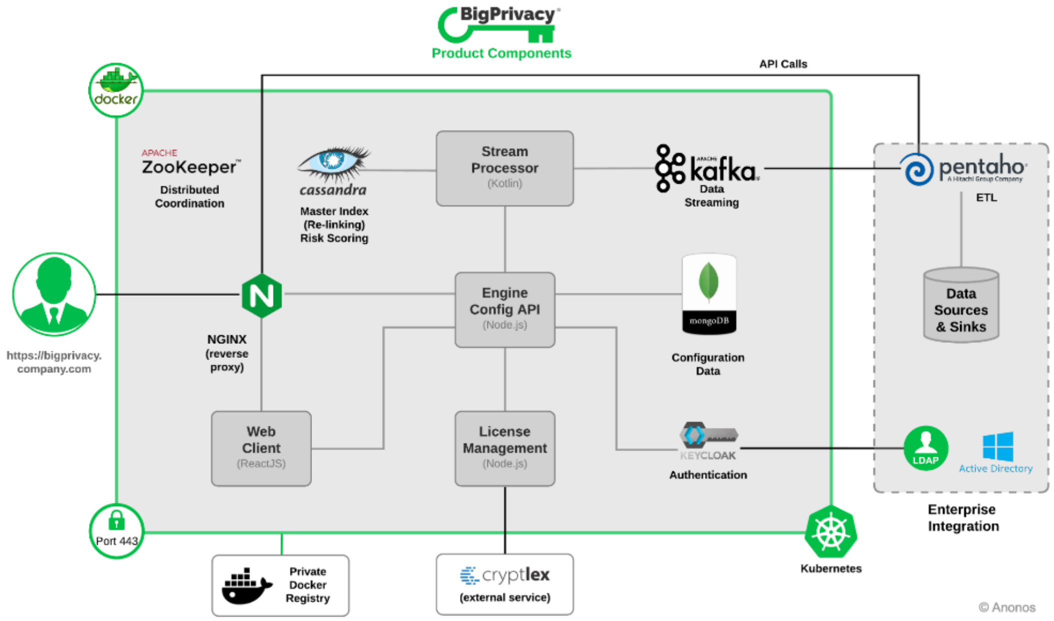
[82] https://www.enisa.europa.eu/publications/recommendations-on-shaping-technology-according-to-gdpr-provisions

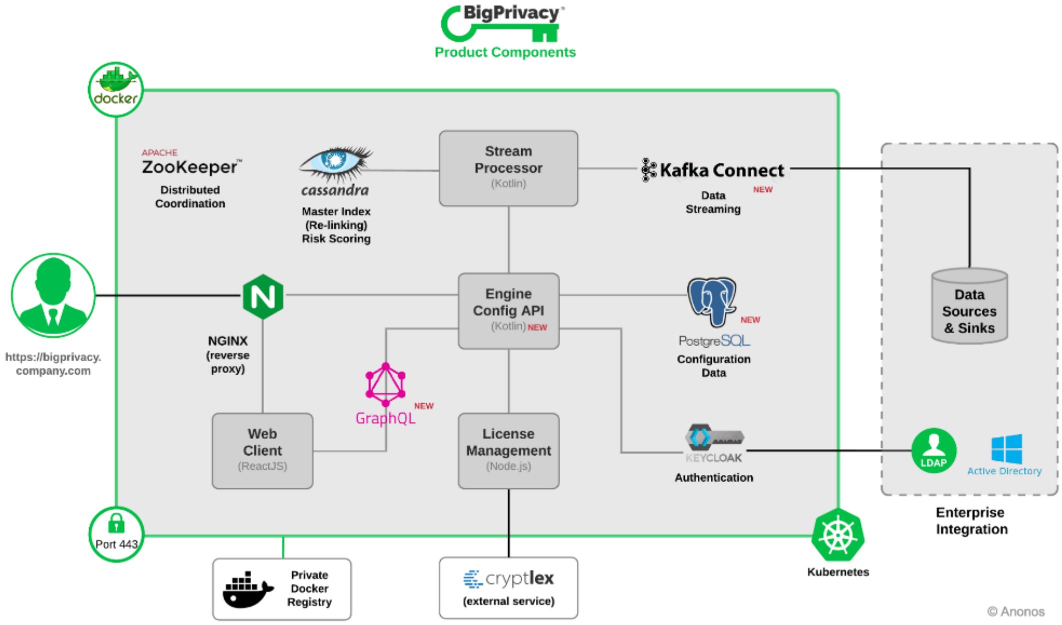
ENISA GUIDELINE NO 2 – NOVEMBER 2019: Recommendations on Shaping Technology According to Data Protection and Privacy Provisions – Pseudonymisation Techniques and Best Practices[83]	**Section**	**Anonos BigPrivacy**
Enables A Risk-Based Approach Accounting for Required Protection and Utility/Scalability	Exec Summary	✓
Advances the State-of-the-Art	Exec Summary	✓
Complies with GDPR Definition of Pseudonymisation	2	✓
Utilizes one or more Pseudonymisation Functions	2	✓
Utilizes a Pseudonymisation Secret	2	✓
Has a Recovery Function for Pseudonymisation Functions	2	✓
Uses a Pseudonymisation Mapping Table	2	✓
Attack Resistance	4.3	✓
Pseudonymisation Secret Discovery Attack Resistant	4.3.1	✓
Re-Identification (Linkage) Attack Resistant	4.3.2	✓
Discrimination (Inference) Attack Resistant	4.3.3	✓
Brute Force Attack Resistant	4.4.1	✓
Dictionary Search Resistant	4.4.2	✓
Utility and Data Protection Maximization	4.5	✓
Pseudonymisation Techniques	5.1	✓
Does Not Make Use of Counters	5.1.1	✓
Uses Cryptographic Random Number Generator	5.1.2	✓
Does Not Use Cryptographic Hash Function with or without Salts, Peppers	5.1.3	✓
Uses MAC – Keyed Hash (HMAC)	5.1.4	✓
Uses Symmetric Encryption	5.1.5	✓
Pseudonymisation Policies	5.2	✓
Supports Deterministic Pseudonymisation	5.2.1	✓
Supports Fully Randomized - RDDIDs - Both Row and Field Level	5.2.3	✓
Offers Recovery Function (Reversal of Pseudonymisation)	5.4	✓
Protects Pseudonymisation Secret	5.5	✓
Advanced Pseudonymisation Techniques	5.6	✓
Controlled Pseudonym Linkability	5.6	✓
K-Anonymity	5.6	✓
Aggregation/Generalization/Binning	5.6	✓
Rounding	5.6	✓
Masking	5.6	✓
Prefix/Suffix-Preserving Pseudonymisation	6.2.1	✓
Format Preserving Pseudonymisation	7.4	✓

[83] https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices

APPENDIX C

BIGPRIVACY ARCHITECTURE





Anonos®

LearnMore@SpeedToInsight.com

www.anonos.com