The GDPR and Controlled Linkable Data

Satisfying GDPR requirements for Anonymisation Pseudonymisation, De-Identification, and Data Protection by Design and by Default

September 2017

How to unlock maximum data value by enabling global compliance controls on a jurisdictional basis

Anonos®

The GDPR and Controlled Linkable Data

Balancing the Interests of Regulators, Data Controllers and Data Subjects

September 2017

Mike Hintze[1]

Gary LaFever[2]

1. Partner at Hintze Law PLLC. Part-time Professor, University of Washington School of Law. Formerly, Chief Privacy Counsel, Microsoft Corporation. This White Paper incorporates elements of my paper "Viewing the GDPR Through a De-Identification Lens: A Tool for Clarification and Compliance" presented at the Brussels Privacy Symposium in November 2016 and available at https://fpf.org/wp-content/uploads/2016/11/M-Hintze-GDPR-Through-the-De-Identification-Lens-31-Oct-2016-002.pdf. Anonos Inc. (Anonos®) is a current client and support for my contribution to this White Paper has been provided by Anonos. Neither my representation of Anonos nor my contribution to this White Paper serves as an endorsement of any technology, products or services of Anonos, including BigPrivacy®, or otherwise. The views expressed in my contribution to this White Paper are my own and do not necessarily reflect the positions of any current or former employer or client. This White Paper is for informational purposes only and is not intended to, nor shall it be construed as, providing any legal opinion or conclusion; does not constitute legal advice; and is not a substitute for obtaining professional legal counsel from a qualified attorney on your specific matter.

2. CEO and Co-Founder at Anonos. Formerly, Partner at Hogan Lovells.

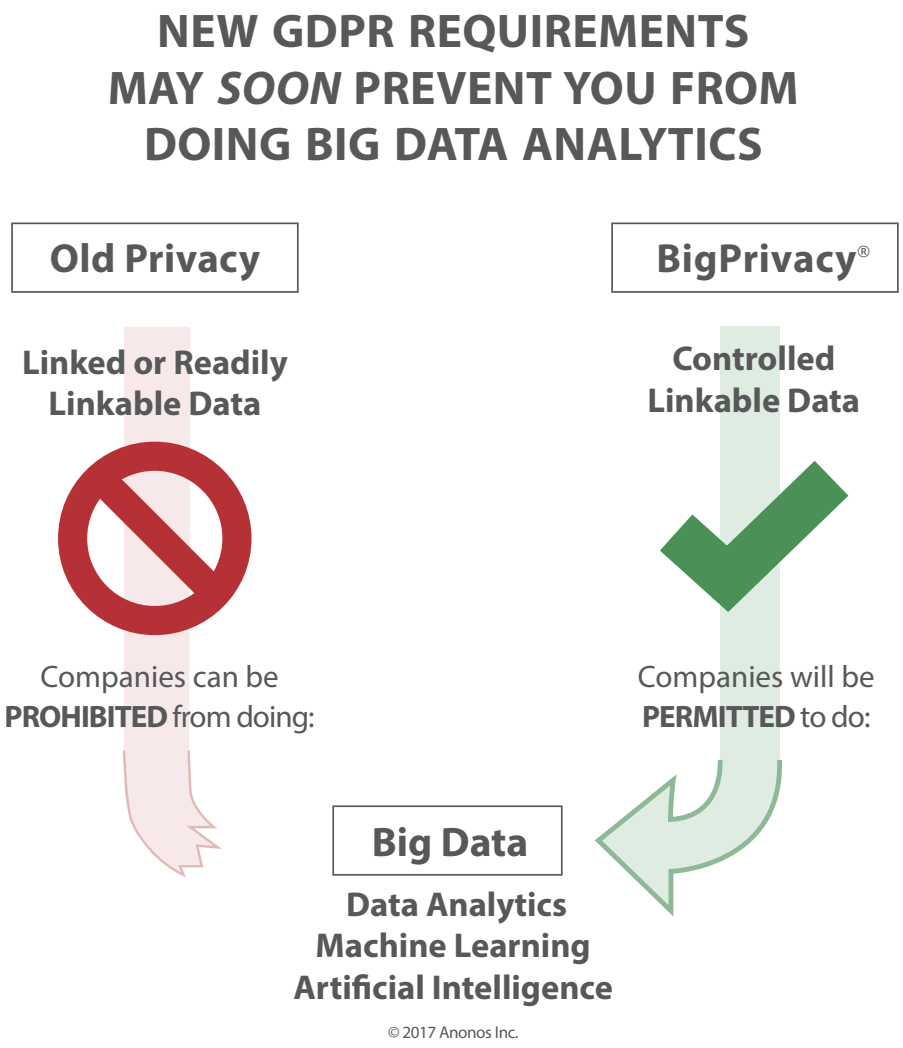
Table of Contents

Appendices

This White Paper was previously published with the title "Meeting Upcoming GDPR Requirements While Maximizing the Full Value of Data Analytics" on the Social Science Resource Network (SSRN) on March 6, 2017 at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2927540. Citations to this SSRN article include, *inter alia*, the following:

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3034261
https://www.cpomagazine.com/2017/08/13/gdpr-locks-data-whats-solution/
https://www.lexology.com/library/detail.aspx?g=f1177675-4fc0-478b-9ae0-5c1f540e91e7
https://www.lexology.com/library/detail.aspx?g=ec568120-5ea4-4509-9412-21121c8c4c9d

I. Infographic



Benefits of Controlled Linkable Data:

Controlled Linkable Data enables intelligent technical and policy solutions that deliver the benefits of data uses while avoiding the risks.

Jules Polonetsky, CEO
Future of Privacy Forum

Controlled Linkable Data tools minimize risk by de-linking and re-linking data to break the stalemate between responsible use and data obscurity.

Martin Abrams, Executive Director & Chief Strategist
Information Accountability Foundation

II. Executive Summary

The new obligations imposed by the General Data Protection Regulation (GDPR) do not prohibit the use of personal data for analytics or other beneficial secondary uses. But they do require the adoption of new technical and organizational measures to protect that data. The GDPR explicitly points to pseudonymizing as one such measure that can help meet the requirements of several of its provisions. The GDPR further recognizes differing levels of de-identification in a way that provides incentives for organizations to adopt the optimal type and level of de-identification that can help them use personal data for beneficial purposes while meeting their compliance obligations and protecting the privacy of individuals.

By enabling the use of "Controlled Linkable Data" (as described in this White Paper) that retains the utility of personal data while helping to meet organizations' compliance obligations and to significantly reduce their risk of liability, Anonos® BigPrivacy® technology can help organizations navigate and meet these new GDPR requirements. Thus, Anonos BigPrivacy technology can ease regulatory burdens and be a key component of an overall GDPR compliance program.

Anonos BigPrivacy technology can ease regulatory burdens and be a key component of an overall GDPR compliance program.

The body of this paper describes in detail the regulatory background, technological innovations, and practical applications of Controlled Linkable Data, leading to the maximization of data value and individual privacy in a GDPR-compliant manner.

First, in Section III, we introduce the concept of Controlled Linkable Data in the context of the GDPR. Next, in Section IV, we describe the GDPR's new requirements, focusing on the distinction between privacy by design and data protection by default, and noting that the former is merely a subset of the latter, making it insufficient to satisfy the GDPR's stringency. We also introduce the essential concept of Controlled Linkable Data. In Section V, we explain how Controlled Linkable Data enables a more powerful form of de-identification, one encouraged by the GDPR, but which has previously not been achievable by technical methods. This leads to the conclusion that "data protection over the full lifecycle of data by leveraging technical and organizational measures, including pseudonymisation, [ensures] that, by default, personal data are not made accessible without the individual's intervention to an indefinite number of natural persons." Next, Section VI analyzes numerous relevant sections of the GDPR (specifically, Articles 5, 6, 11(2), 12(2), 15-22, 32-36, 40, 42, 82 and 88), showing how Controlled Linkable Data helps satisfy the specific GDPR requirements. Last, in light of this understanding of the requirements, limitations, exclusions and overall principles of the GDPR, Section VII explains the technical basis of Anonos BigPrivacy technology, how it implements Controlled Linkable Data, and how this solution addresses GDPR compliance concerns for all parties: data controllers, regulators and data subjects.

Global firms that gather, use or store GDPR personal data should consider the possibility that Controlled Linkable Data as described in this White Paper enables secondary uses of data while ensuring compliance with GDPR requirements.

III. Controlled Linkable Data and the GDPR

Data protection law in the European Union (EU) and the European Economic Area (EEA) is about to radically change, directly impacting every global organization which handles personal data pertaining to any EU/EEA person and even exposing such organizations to the risk of *penalties as high as 4 percent of their worldwide gross revenues.*

Under the GDPR, which will assume the force of law on May 25, 2018, these organizations will not only have to comply with far stricter data protection obligations and the establishment of new rights for individual data subjects, but they will also have to adopt *unprecedented new technical measures* that enable such compliance and enforce these rights. While it is true that a cloud of uncertainty remains about precisely how certain GDPR regulations may be applied, there is no doubt whatsoever that organizations will have to adopt *unprecedented* new measures to establish what the GDPR terms "data protection by default." This is theoretically intriguing, but, as a practical matter, potentially devastating.

The reason is that the onset of the GDPR regime creates a significant dilemma for all global data controllers and processors: either comply with the GDPR and its "data protection by default," but endure significant limits on data use and data value – or subject oneself to the possibility of debilitating fines (yearly, multi-billion-euro fines for the largest companies).

While it is true that a cloud of uncertainty remains about precisely how certain GDPR regulations may be applied, there is no doubt whatsoever that organizations will have to adopt *unprecedented* new measures to establish what the GDPR terms "data protection by default."

This paper describes how a new technical approach to de-identification – "Controlled Linkable Data" – allows data use and the unlocking of data value in a way that enables compliance with the GDPR, all while enhancing individual data subject privacy. In brief, Controlled Linkable Data represents a potential cornerstone technological approach for data controllers, regulators and data subjects for three key reasons:

1. It decouples data elements from re-identifiable linkages to data subjects while enabling selected portions of that data (or abstracted ranges or groupings of that data) to be available to legitimate, authorized users for specific times or purposes – or in specific places – in a way that can be fully GDPR-compliant (this means providing technical and organizational measures to enable GDPR-compliant secondary uses of data like analytics, machine learning and artificial intelligence);

2. It defines, enables and manages multiple, different levels of de-identification, based on situation-specific applicable regulatory and policy implications, producing the highest practical level of de-identification for any given data use, and optimizing the balance between maintaining the utility of data and protecting privacy and security; and

3. It enables data controllers to meet the precise *exclusionary* standards of GDPR Articles 11(2) and 12(2), because the de-identification performed uses unique technological means to sever the re-identification links between data elements and data subjects; this renders data controllers without re-identification key access (cf. discussion on JITI® keys in this paper) "not in a position to identify data subjects," thus meeting the exclusion criterion set forth by the GDPR.

IV. Analytics and Secondary Use of Data Under the GDPR

Analytics, machine learning, and artificial intelligence (AI), together with other secondary uses[3] of data, are at the cornerstone of the new global digital economy. However, industry experts, privacy lawyers, and data protection professionals have raised important concerns about the compatibility of such data uses with the heightened obligations for protecting personal data under the new GDPR.

Some have gone so far as to conclude that analytical and other secondary uses of data are impractical, if not impossible or illegal, under the GDPR.

Penalties of up to 4 percent of offenders' worldwide gross revenues starting on May 25, 2018, and far-reaching joint and several liability obligations among data controllers and processors under GDPR exponentially exacerbate the severity of these concerns. Some have gone so far as to conclude that analytical and other secondary uses of data are impractical, if not impossible or illegal, under the GDPR.

For the first time, the GDPR specifically requires privacy by design under data protection legislation.[4] However, the GDPR requires more than *just* privacy by design; it requires data protection by default, the most stringent implementation of privacy by design. As more fully described below, the GDPR now requires that data protection by default be applied at the earliest opportunity (e.g., by pseudonymizing data at the earliest opportunity) to limit data use to the minimum *extent* and *time* necessary to support each specific product or service authorized by an individual data subject.

A paper entitled *Viewing the GDPR Through a De-Identification Lens: A Tool for Clarification, Compliance, and Consistency* was presented by Mike Hintze at the November 8, 2016, Brussels Privacy Symposium on Identifiability: Policy and Practical Solutions for Anonymization and Pseudonymization at the Vrije Universiteit, Brussels (the "Brussels GDPR De-Id Paper"). This White Paper builds upon the Brussels GDPR De-Id Paper to highlight how complying with GDPR data protection by default obligations by de-identifying data can bridge the gap between GDPR's obligations and secondary uses of data.

A data protection by default approach to de-identification leverages incentives built into the GDPR to use technical measures to enable GDPR-compliant secondary use of data. Traditional technologies like encryption, hashing, and Privacy Enhancing Techniques (PETs, e.g., k-anonymity, l-diversity, and differential privacy) were developed long before GDPR requirements were established. When used alone, encryption, hashing, and PETs may fail to satisfy the GDPR's data protection by default requirements.

---

3. Secondary uses of data are uses other than those that undergird the initial creation or capture of data. For example, data uses necessary to ensure that an employee is paid and that payment is deposited at the employee's bank would constitute primary use. Similarly, data uses necessary to ensure that a telephone call is connected, or an email is transmitted, to an intended recipient constitute primary data use. Conversely, use of employee-related data to compare an employee to other parties and use of telephone/email data to make an inference about the initiator or recipient of information from a call/email constitute secondary uses of data.

4. Privacy by Design is the approach championed by Ann Cavoukian, Ph.D., former Information and Privacy Commissioner of Ontario, for embedding privacy into the system design process. See https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf.

Static "anonymous/pseudonymous" tokens[5] may fall short because links between data subjects and identifying information remain readily ascertainable. Consequently, European regulators may conclude that these techniques fail to satisfy data protection by default requirements due to re-identification risks from linkage attacks and the Mosaic Effect.[6] Finally, stateless "anonymous/pseudonymous" tokens[7] developed for PCI compliance in the payment card industry fail to enforce re-linking and revealing of personal data under the controlled conditions necessary to support iterative analytics and secondary uses of data.

The GDPR provides incentives to use technical measures that enable the flow, commercial use, and value maximization of data in a way that recognizes, respects, and enforces the rights of individuals. This may be accomplished by controlling the linkability of identifying information to individual data subjects.

This White Paper highlights unique capabilities of controlled linkability technology from Anonos called "BigPrivacy." Controlled linkability leverages technical and organizational measures to support a wide spectrum of de-identification, with the different levels of de-identification having different regulatory and policy implications. By enforcing the highest practical level of de-identification for a given data use, controlled linkability helps to achieve the optimal balance between the utility of data and protecting privacy and security.

Controlled linkability achieves a win-win-win for regulators, data controllers, and data subjects alike.

Controlled linkability achieves a win-win-win for regulators, data controllers, and data subjects alike. As more fully described below, controlled linkability helps data controllers and processors with specific GDPR obligations (e.g. data protection by default, purpose limitation, data minimization, data breach notification, etc.). It also provides specific regulatory and compliance benefits to recipients of data who are not in control of keys necessary to re-identify data (i.e., relief under Articles 11(2) and 12(2) from certain data subject obligations). Further, the benefits of controlled linkability from a data protection by default approach to de-identification as outlined in this White Paper extend beyond GDPR compliance to enable controls necessary for secondary uses of data underlying the new global digital economy.

V. Controlled Linkable Data in Support of De-identification

De-identification techniques help to reduce privacy risks and protect data subjects' rights under the GDPR. As noted in the Brussels GDPR De-Id Paper, the GDPR recognizes a spectrum of de-identification. Anonymization, requiring irreversible de-linking of data from data subjects so that re-identification of data subjects is no longer possible, represents the far end of the de-identification spectrum. This

5. Static "anonymous/pseudonymous" tokens are tokens used on a consistent or persistent basis to replace identifying information.

6. The "Mosaic Effect" occurs when it is possible to determine an individual data subject's identity without having access to primary identifiers (e.g., name, data of birth, street address, etc.) by correlating data pertaining to the individual across numerous data sets.

7. Stateless "anonymous/pseudonymous" tokens are tokens that change frequently to replace identifying information.

recognition of a de-identification spectrum is an important departure from the largely binary approach European regulators have taken to date – with data being classified as either personal data and therefore subject to data protection law, or as anonymous and therefore not subject to data protection law. This binary approach often fails to provide the necessary incentives to use the most robust de-identification compatible with the intended data uses, and it therefore leads to suboptimal results. In contrast to this binary approach to de-identification, Anonos BigPrivacy technology supports a dynamic, technology-enforced, flexible approach to de-identification referred to herein as "Controlled Linkability" or Anonosizing® data.

With Controlled Linkability, compliance with GDPR and other data protection/restriction obligations is facilitated because data can be pseudonymized to meet certain GDPR requirements such as data protection by default. Moreover, the same data can represent one level of de-identified data to one entity, and another level to another entity – depending on who controls keys necessary to re-identify the data. In this manner, the level of identifiability is related to the entity processing the data. For the entity holding re-identification keys, the data may represent one level of de-identification, but for an entity that does not control or have access to the keys, the data may represent a higher level of de-identification. Exclusions under GDPR Articles 11(2) and 12(2) for controllers "not in a position to identify data subjects" may apply to controllers who do not have access to those keys, but they may not apply to controllers with access to both data and the keys necessary to re-identify it.

> De-identification techniques help to reduce privacy risks and protect data subjects' rights under the GDPR.

De-identification is a process that can be used to remove or obscure personal information from data to enable use, storage, and sharing of data. In 2015, the Information Access Division of the U.S. National Institute of Standards and Technology published a report on de-identification, "NISTIR 8053: De-identification of Personal Data" (the "NIST De-Id Report"). Section 2.5 of the NIST De-Id Report provided a summary of three prevalent de-identification models:

1. **The Release and Forget model:** The de-identified data may be released to the public, typically by being published on the Internet. It can be difficult or impossible for an organization to recall the data once released in this fashion.

2. **The Data Use Agreement model:** The de-identified data may be made available under a legally binding data use agreement (DUA) that details what can and cannot be done with the data. Typically, data use agreements prohibit attempted re-identification, linking to other data, or redistribution of the data. DUAs will typically be negotiated between the data holder and qualified researchers (the "qualified investigator model"), although they may be simply posted on the Internet with a click-through license agreement that must be agreed to before the data can be downloaded (the "click-through model").

3. **The Enclave model:** The de-identified data may be kept in a segregated enclave that restricts the export of the original data, and instead accepts queries from qualified researchers, runs the queries on the de-identified data, and responds with results.

In effect, the GDPR introduces a fourth model of de-identification: **Data Protection by Default,** which combines elements of the three models above. Recital 78 of the GDPR describes data protection

by default as including "…pseudonymising personal data as soon as possible." Article 4(5) defines pseudonymisation as "processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, ***provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person***" (emphasis added). In addition, Article 25, Data Protection by Design and by Default, states:

1. Taking into account the ***state of the art***, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, ***the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organizational measures, such as pseudonymisation,*** which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.

Viewed as a fourth model of de-identification, data protection by default supports data protection over the full lifecycle of data by leveraging technical and organizational measures, including pseudonymization, to ensure that, by default, personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.

2. ***The controller shall implement appropriate technical and organizational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed.*** That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, ***such measures shall ensure that, by default, personal data are not made accessible without the individual's intervention to an indefinite number of natural persons (emphasis added)***.

Viewed as a fourth model of de-identification, data protection by default supports data protection over the full lifecycle of data by leveraging technical and organizational measures, including pseudonymization, to ensure that, by default, personal data are not made accessible without the individual's intervention to an indefinite number of natural persons. By providing a clear roadmap for compliance, while incentivizing the optimal level of de-identification to both enable beneficial data use and protect individual privacy, data protection by default enables a win-win-win for regulators, data controllers, and data subjects alike.

To highlight de-identification attributes specific to Controlled Linkable Data, this White Paper uses the following terminology and taxonomy rather than a more generalized terminology and taxonomy such as presented in the Brussels GDPR De-Id Paper. *A detailed explanation of how the technological tools described here support different levels of de-identification is provided in Section VII.*

1. **Linked Data** - Identifies or is directly linked to data that identifies a specific natural person (such as a name, e-mail address, or government-issued ID number).

2. **Readily Linkable Data** - Relates to a specific person whose identity is not apparent from the data; the data is not directly linked with data that identifies the person, but there are readily available means to create or re-create a link with identifying data. Data that fails to satisfy the requirements for GDPR Article 4(5) compliant pseudonymized data (i.e., "information necessary to attribute personal data to a specific data subject must be kept separately and subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person") is a subset of Readily Linkable Data. For example, data that uses static, so-called "anonymous/pseudonymous" identifiers, without other protections, constitute Readily Linkable Data because they fail to protect against privacy breaches from linkage attacks and re-identification via the Mosaic Effect.

3. **Controlled Linkable Data** - Relates to a specific person whose identity is not apparent from the data; the data is not directly linked with data that identifies the person, and effective technical and organizational measures are in place controlling access to create or re-create a link with identifying data. Anonos BigPrivacy supports Controlled Linkable Data by using different Dynamic De-Identifiers® (DDIDs®) to represent data elements and requiring the use of Just-In-Time-Information™ (JITI) keys to re-identify or re-associate data.

    a. Privacy Rights Management for De-Identification™ or PRMD™ (as more fully described in Section VII below): JITI keys are controlled by data controllers and/or data processors to satisfy requirements for GDPR Article 4(5)-compliant pseudonymized data (i.e., "information necessary to attribute personal data to a specific data subject must be kept separately and subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person").

    b. Privacy Rights Management for Individuals™ or PRMI™ (as more fully described in Section VII below): When JITI keys revealing non-identifying information are controlled by data controllers and/or data processors in a manner that satisfies requirements for GDPR Article 4(5)-compliant pseudonymized data (i.e., "information necessary to attribute personal data to a specific data subject must be kept separately and subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person").

4. **Anonymous Data** - To achieve anonymity so data does not fall within the scope of the GDPR, certain stringent conditions must be met.[8] For Privacy Rights Management for Individuals (PRMI) (described below) to support anonymity, adequate data policies must be in place (e.g.,

8. GDPR Recital 26 stipulates that "The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken 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. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes."

policies must prevent the public release of a de-identified data set that could be matched to other publicly available data to re-identify and the nature of the protected data must not be inherently identified). For Anonymization/Aggregation (described below) to support anonymity, methods must be irreversible and eliminate any known or foreseeable possibility of linking any of the data to an individual to whom the data originally related.

a. Privacy Rights Management for Individuals or PRMI (as more fully described in Section VII below): When JITI keys necessary to reveal identifying information are controlled by data subjects or by trusted third parties specifically authorized by data subjects to satisfy:

   i. GDPR Recital 26 requirements that, in determining that data is anonymous, account should be taken of (a) all 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, (b) whether means are reasonably likely to be used to identify the natural person, taking into account all objective factors, such as the costs of and the amount of time required for identification, (c) available technology at the time of the processing, and (d) technological development; and

   ii. WP29 Opinion 05-2014 on Anonymisation Techniques[9] requirements that (a) anonymous data must be processed in such a way that it can no longer be used to identify a natural person by using "all the means likely reasonably to be used" by either the controller or a third party, and (b) WP29 clarification that the "means ... reasonably to be used" test is suggested by the Directive as a criterion to be applied in order to assess whether the anonymisation process is sufficiently robust, i.e. whether identification has become "reasonably" impossible.

b. Anonymous/Aggregate Data: Data that is stored without any identifiers or other data that could identify the individual or device to whom the data relates and aggregated with data about enough individuals such that it does not contain individual-level entries or events linkable to a specific person.

VI. Benefits of Processing Controlled Linkable Data

1. Controlled Linkable Data helps satisfy GDPR Article 5 requirements that processing of personal data comply with the following:

   a. **Purpose Limitation**: Personal data should be collected for specified, explicit, and legitimate purposes and not further processed in a manner incompatible with those purposes.

      i. Further processing for statistical purposes is not considered incompatible with the Purpose Limitation if such processing is subject to Article 89.1 safeguards for the rights and freedoms of data subjects including that "safeguards shall ensure that technical and organizational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided

9. Article 29 Data Protection Working Party Opinion 05/2014 on Anonymization Techniques, WP216, adopted 10 April 2014. Note that while this Working Party opinion interprets the 1995 Data Protection Directive, which is repealed and replaced by the GDPR, the concept of anonymization is largely unchanged by the GDPR and the Working Party's analysis is therefore of continued relevance.

that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner."

b. **Data Minimization**: Data processing should be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed.

c. **Storage Limitation**: Personal data is to be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed. Personal data may be stored for longer periods if processed solely for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organizational measures to safeguard the rights and freedoms of data subjects.

d. **Integrity and Confidentiality**: Data is to be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction, or damage, using appropriate technical or organizational measures.

2. Controlled Linkable Data helps to satisfy GDPR Article 6 requirements for the lawful processing of personal data. As noted in the Brussels GDPR De-Id Paper, the GDPR makes it more difficult to rely on consent of a data subject as the sole legal basis for processing. Technological advances such as the Internet of Things, big data analytics, machine learning, and artificial intelligence (AI) make reliance on consent impractical in some instances. In addition, certain categories of data involve an imbalance of power between data subjects and data controllers (e.g., employment-related data in the context of employees and employers), thereby requiring additional legal bases for processing beyond consent alone.

   a. **Legitimate Interest:** Article 6(4) of the GDPR supports the idea that de-identification may undergird "legitimate interest" as a legal basis for processing data. "Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent . . . the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, *inter alia* . . . (e) the existence of appropriate safeguards, which may include encryption or pseudonymisation." Controlled Linkable Data helps to enforce these safeguards.

3. The processing of Controlled Linkable Data helps data controllers benefit from provisions under GDPR Articles 11(2) and 12(2). Controllers that do not control or have access to the re-identification keys (e.g. the JITI keys) are "not in a position to identify the data subject." Thus, under Article 12(2), the data controller is not subject to the following rights of data subjects:

   a. Article 15 - Right of Access

   b. Article 16 - Right to Rectification

   c. Article 17 - Right to Erasure/Right to be Forgotten

   d. Article 18 - Right to Restrict Processing

e. Article 19 - Notification to Data Recipients of any Rectification, Erasure, or Restriction of Processing

f. Article 20 - Data Portability

g. Article 21 - Right to Object

h. Article 22 - Exclusion from Automated Decision-Making/Profiling

4. Even for those controllers that have access to the re-identification keys and thus are "in a position to identify the data subject" such that the exclusions of Articles 11(2) and 12(2) would not apply, the use of Controlled Linkable Data can nevertheless help such controllers comply with applicable data subject rights. For example, Controlled Linkable Data may be used to enable compliance with data subjects' Right to Erasure/Right to be Forgotten under Article 17. As highlighted in Appendix C, Anonos BigPrivacy technology can control removal, restoration, and protection of linkages between and among data sets and known identifiers. As such, it can enable the permanent destruction of links so that they cannot be recovered. This permanent "orphaning" of data, such that it cannot be re-linked, can be viewed as equivalent to deletion – in much the same way that encrypting and throwing away the key are viewed as equivalent to deletion.

5. The processing of Controlled Linkable Data can constitute technical and organizational measures that safeguard rights of data subjects. Such measures are helpful in complying with additional GDPR obligations, *inter alia*, under:

   a. Article 32 - Security of Processing.

      With respect to data security, the GDPR specifies that both data controllers and data processers must "[t]aking into account the state of the art, ... implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including *inter alia* as appropriate ... the pseudonymisation and encryption of personal data ...." Using Controlled Linkable Data can thus constitute a key element of the required security measures.

   b. Article 33 - Notification of Personal Data Breach to Supervisory Authority.

      Under the GDPR, the obligation to provide notification of a data breach is tied to the likelihood of risk to the rights and freedoms of natural persons. Specifically, controllers must notify the supervisory authority "unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons." The use of Controlled Linkable Data has a direct bearing on that risk, and can help data controllers determine their obligations under Article 33.

   c. Article 34 - Communication of Personal Data Breach to Data Subjects.

      As with the requirements under Article 33, the Article 34 obligation to notify data subjects of a data breach hinges on the likelihood of risk to the rights and freedoms of natural persons. In this case, the obligation applies when that risk is "high." Here, too, the use of Controlled

Linkable Data will ease that determination and reduce the risk to individual data subjects' rights and freedoms.

d. Article 35 - Data Protection Impact Assessments (including, *inter alia,* processing of special categories of data (e.g., health data) identified in Article 9(1)).

   Whether a Data Protection Impact Assessment (DPIA) is required under the GDPR also depends upon whether the processing is likely to result in a high risk to the rights and freedoms of natural persons. In cases where a DPIA is required (such as the processing of special categories of sensitive data), it must include, among other things, "the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation." The use of Controlled Linkable Data can reduce the need for a DPIA in some cases, and where one is needed, it can be a key part of the "measures" that must be documented.

e. Article 36 - Prior Consultation Obligations.

   Where a DPIA indicates that data processing would result in a high risk in the absence of mitigation steps, controllers are obligated to consult with the supervisory authority prior to processing the data. As part of the consultation, the controller must provide the DPIA with a description of the "measures and safeguards provided to protect the rights and freedoms of data subjects pursuant to [the] Regulation." Controlled Linkable Data can help satisfy the requirement for "measures and safeguards" that protect the rights and freedoms of data subjects.

f. Article 82 - Rights to Compensation and Liability.

   The processing of Controlled Linkable Data can alleviate the burden of joint and several liability among data controllers and processors by helping them comply with obligations under the GDPR and reducing the likelihood of damage.

6. Documentation evidencing technical and organizational measures made possible by using Controlled Linkable Data are helpful in establishing the standard of care under the GDPR including, *inter alia*, the following:

   a. Article 40 - Establishing Codes of Conduct

   b. Article 42 - Establishing Certification Mechanisms/Data Protection Seals/Marks

   c. Article 88 - Processing in the Context of Employment

VII. Controlled Linkability with Anonos BigPrivacy Technology

Anonos BigPrivacy technology[10] minimizes unauthorized re-identification by "Anonosizing" data, a patented process that infuses granular privacy and security controls into data at scale. Anonosizing

10. Anonos BigPrivacy technology is protected by an intellectual property portfolio that includes four granted U.S. patents (9,361,481; 9,129,133; 9,087,216; and 9,087,215) and 50+ additional U.S. and international patent applications. Anonos, Anonosizing, BigPrivacy, Circle of Trust, CoT, DDID, Dynamic De-Identifier, JITI, Just-In-Time-Information, Privacy Rights Management, Privacy Rights Management for De-Identification, Privacy Rights Management for Individuals, PRM, PRMD, and PRMI are trademarks of Anonos protected under domestic and international laws and treaties.

takes Digital Rights Management (DRM) techniques like those used by companies to limit copies that individuals can make of music, movies, and other digital content and – uniquely – "stands DRM on its head" as described in the TED Talk noted below.[11] BigPrivacy shifts the power from the corporate owner to the data subject by enabling a data subject, or an entity that a data subject trusts, to authorize uses of the data subject's personal data. This is called Privacy Rights Management™ or PRM.™ Even in situations where data subjects are not directly involved, Anonos BigPrivacy technology manages risk to enable responsible use of data that respects the rights of data subjects.

> Anonos BigPrivacy technology minimizes unauthorized re-identification by "Anonosizing" data, a patented process that infuses granular privacy and security controls into data at scale.

PRM replaces static, ostensibly anonymous identifiers with dynamically changing, highly protective identifiers (each referred to as a Dynamic De-Identifier or DDID). These dynamic identifiers encapsulate data and provide control over re-identification, throughout the full lifecycle of data, down to the data element level. Thus, the same data can mean different things to different people based on technologically enforced policy controls.

Anonos BigPrivacy technology separates sensitive or identifying data into segments and dereferences these segments using DDID pointers that obscure identities of, and relationships between and among, segmented data elements. Privacy, security, and protection of data are thereby improved by dynamically controlling levels of de-identification.

BigPrivacy technology dynamically controls gradations of data obscurity at different times, for different purposes, at different places and/or by different users – implementing and enforcing policies technologically. These gradations can also impose common data schemata on data collected from different applications and/or platforms thereby enabling functional interoperability among heterogeneous data sets to support data fusion, big data analytics, machine learning and artificial intelligence (AI). Anonosized data is decoded under controlled conditions to support certain uses within designated contexts as authorized by a data subject or by an authorized third party (Trusted Party).

BigPrivacy technology retains the full capability of reproducing up to 100 percent of the original value and utility of data, but it only authorizes the level of identifying information necessary to support each designated use. BigPrivacy controls "identifying" and "associating" data elements so data uses are isolated to those properly permissioned by means of keys/schemata (Just-In-Time-Information or JITI keys/schemata – referred to as "JITI keys") that provide context and meaning for DDIDs. If new authorized data uses arise, all original data value and utility are retained to support them.

Anonos BigPrivacy transforms data by leveraging:

1. **Selectivity** - data elements containing sensitive or identifying information are selected.
2. **Dereferencing** - sensitive data elements are replaced by pseudonymous tokens that serve as

11. Ted Myerson, Co-Founder of Anonos, presented a TED Talk on how BigPrivacy technology enforces Privacy Rights Management or PRM by "standing DRM on its head." A video of, and the transcript for, this TED Talk is at https://anonos.com/TEDTalk. TED Talks is a trademark of Ted Conferences, LLC.

pointers to enable controlled access to data represented by the tokens under technologically enforced conditions.

3. **Dynamism** - BigPrivacy assigns different pseudonymous tokens at different times, for different purposes, at different places, for different users, etc. For example, it has been proven that correlating birthdate, gender, and zip code values from three so-called "anonymous" databases results in the ability to re-identify up to 62 percent of the people in the United States *by name*. However, to combine birthdates, gender, and zip codes to achieve this re-identification, *data must be known a priori to relate to the same individual.* By associating a different pseudonymous token with each person in each of the three databases, it is not evident when birthdates, genders, and zip codes relate to the same person.

Each BigPrivacy pointer is comprised of a dynamically changing[12] DDID. In this manner (i) identities of segmented DDID pointers and (ii) associations between and among segmented DDID pointers are not evident without access to JITI keys to provide context and meaning for each DDID pointer.

The original value of each designated primary or indirect (quasi) identifier can be replaced with two types of DDIDs depending on the type of dereferencing:

1. **Identity Dereferencing** - where a DDID used to replace a data element is intended to point to the value of the replaced data element via a JITI key, the DDID is referred to as a "Replacement DDID" or "R-DDID."

2. **Association Dereferencing** - where a DDID used to replace a data element is intended to both (i) point to the value of the replaced data element via JITI keys; and (ii) convey a range or other association/correlation of the replaced data element to impart information value in a non-identifying manner, the DDID is referred to as an "Association DDID" or "A-DDID."

R-DDIDs and A-DDIDs change dynamically and are temporally unique[13] when used for a different analysis or purpose.

BigPrivacy provides localized, technology-enforced policies for controlling sharing of "Anonosized" data in a dynamically de-identified format. Access to perturbed or original versions of data is controlled by each data source via JITI keys that are under each source's control.

The following figures and appendices provide additional details regarding Anonos BigPrivacy technology.

12. The term "dynamically changing" means that a DDID assigned with respect to a data element representing a data subject, action, activity, process, or trait: (a) changes over time due to (i) passage of a predetermined amount of time, (ii) passage of a flexible amount of time, (iii) expiration of the purpose for which the DDID was created, or (iv) a change in the virtual or real-world location associated with the data subject, action, activity, process, or trait; or (b) is different at different times (i.e., the same DDID is not used at different times) with respect to a same or similar data subject, action, activity, process, or trait.

13. The phrase "temporally unique" means that the time of initial assignment of a DDID to a data subject, action, activity, process, or trait is known, but the time period of assignment may be of any duration, from limited to perpetual.

Figure 1 highlights capabilities of Anonos BigPrivacy for ingesting different data sources and types:

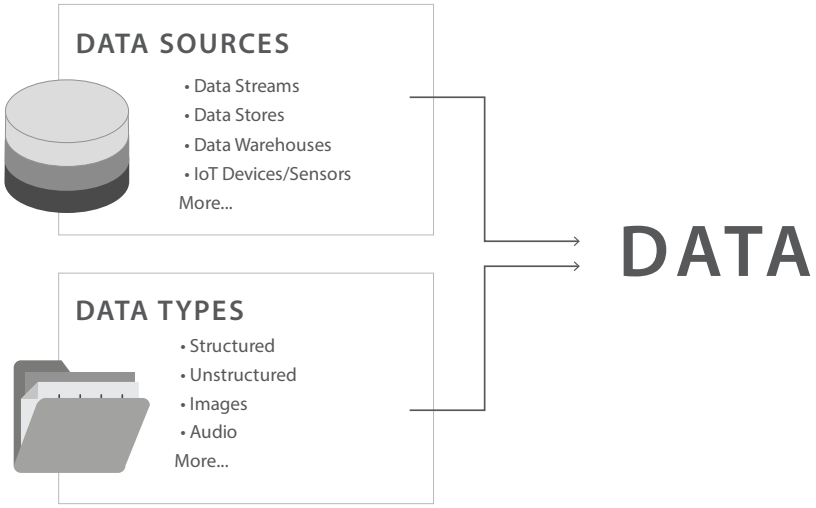


Figure 1

Figure 2 shows how Anonos BigPrivacy technology inserts dynamic de-identification into data sets by transforming direct and indirect identifiers into Replacement De-Identifiers (R-DDIDs) and Association De-Identifiers (A-DDIDs). This dynamism helps reduce risk of re-identification via linkage attacks and the Mosaic Effect.

See https://anonos.com/unicity for an interactive example of dynamic de-identification.

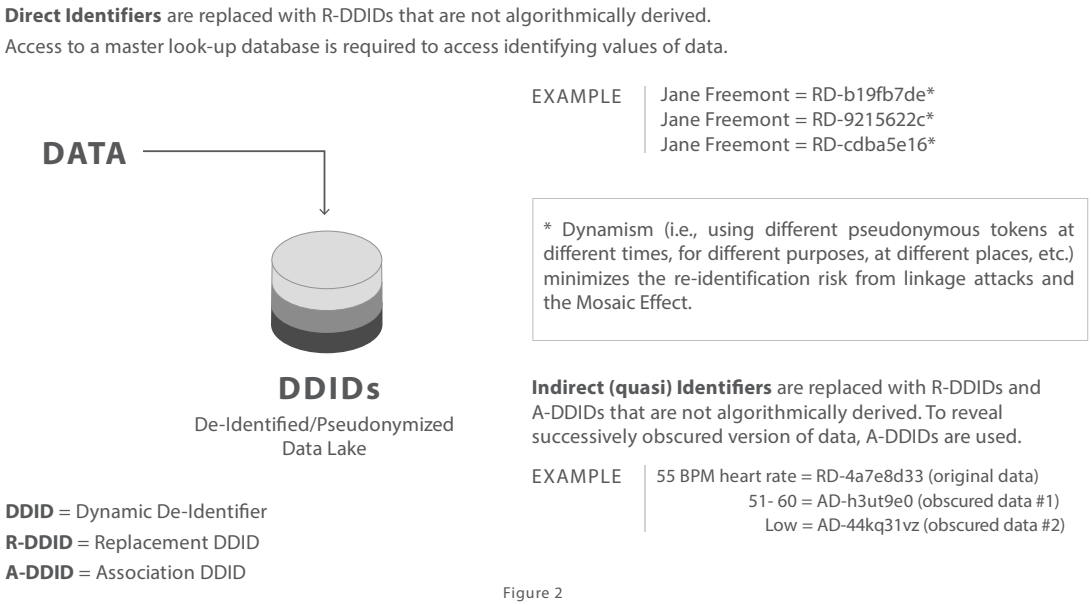


Figure 2

Figure 3 shows how a BigPrivacy-enabled secure environment (i.e., an Anonos enabled Circle of Trust® or CoT®) leverages a master look-up database to support Privacy Rights Management (PRM) for De-Identification (PRMD) and for Individuals (PRMI).

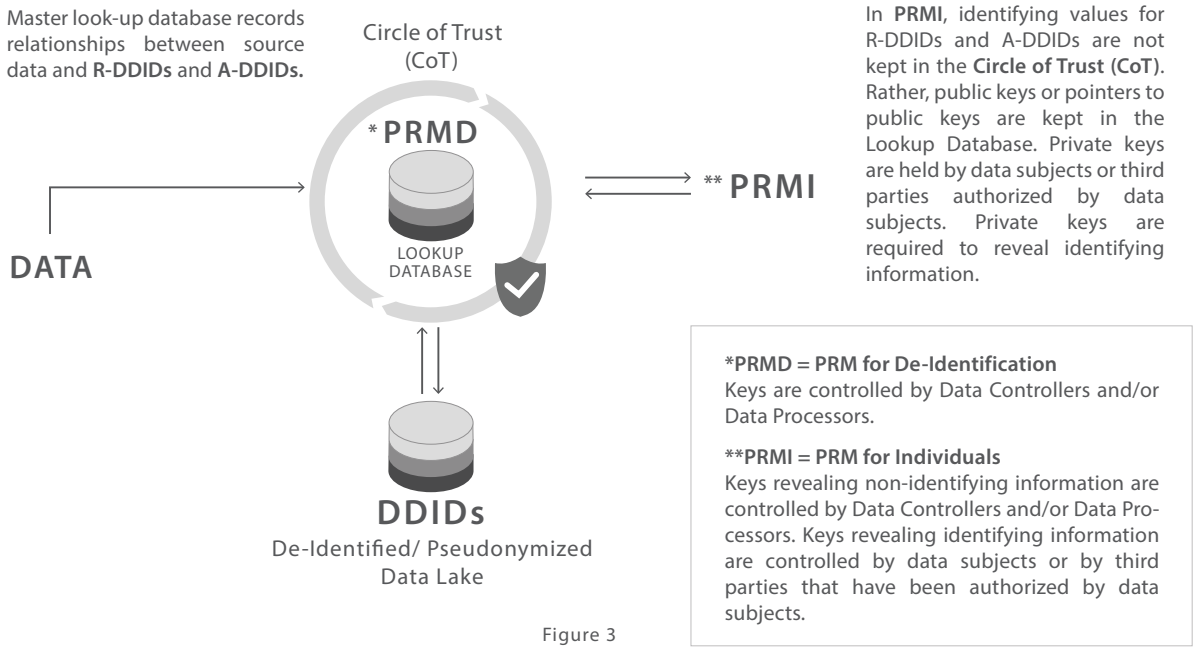


Figure 3

Figure 4 highlights how BigPrivacy controls re-identification by means of Just-In-Time-Information (JITI) keys to selectively reveal data values to which R-DDIDs and A-DDIDs point.

See https://anonos.com/widget for an interactive example of controlled re-identification.

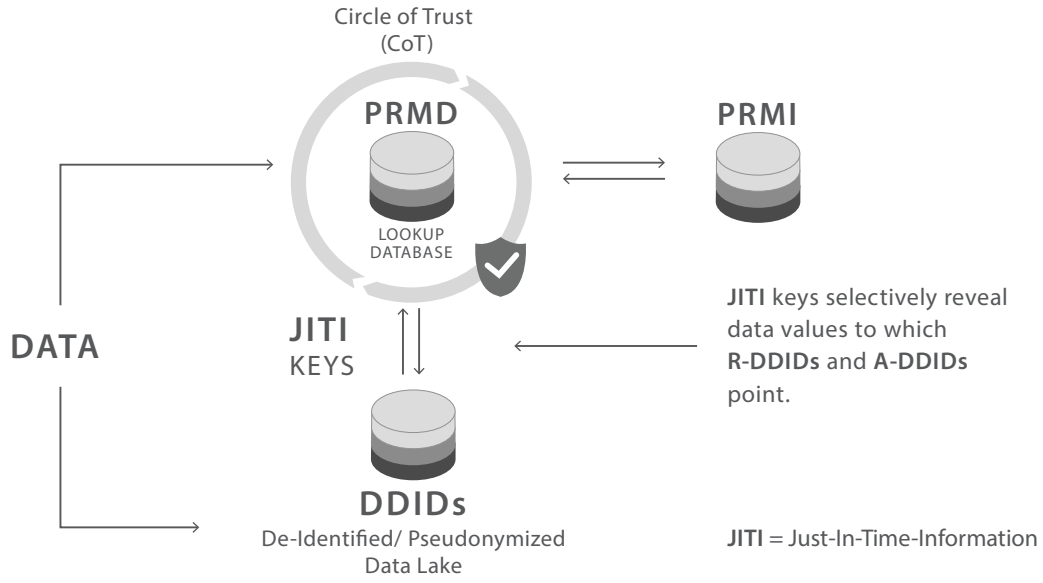


Figure 4

Figure 5 below provides details on an Anonos enabled “Circle of Trust” or “CoT.”

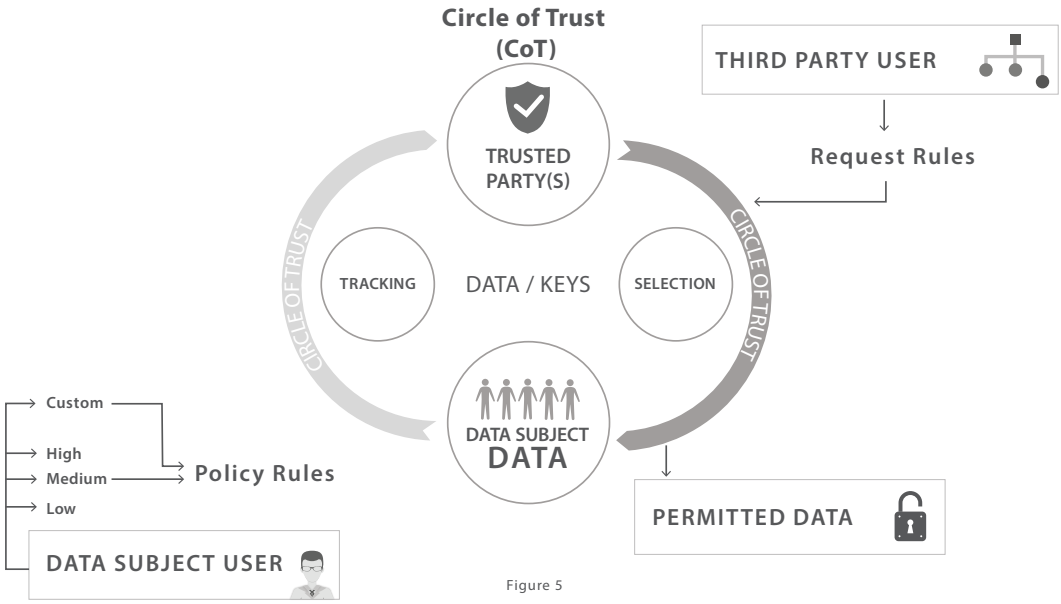


Figure 5

- **Tracking** = Tracking DDIDs used at different times for different purposes.
- **Selection** = Selecting JITI keys necessary to correlate information represented by DDIDs for authorized purposes.
- **Policy Rules** = Allowable operations such as which data can be used by whom, for what purpose, over what time period, etc. Policy Rules may also specify desired “Anonosizing” levels such as when/where/how to use DDIDs for dynamic obscuring in the context of providing protection for the identity and/or activities of a data subject, when to use Privacy Enhancing Technologies (PETs) in connection with DDIDs, when to provide identifying information to facilitate transactions, etc.
- **Third Party User** = A party who inputs data into the system other than the data subject to whom such data relates.
- **Request Rules** = Rules that prescribe data use/access in compliance with established corporate, legislative and/or regulatory data use/privacy requirements.
- **Permitted Data** = Data available for sharing with parties external to the CoT that satisfy Policy Rules established by the Data Subject User and/or Request Rules established by a Third Party User.
- There can be more than one Trusted Party authorized to work within a single Anonos-enabled CoT and data subjects may participate in an unlimited number of CoTs. For increased security, CoTs can be implemented by means of centralized or federated models. Data may remain in original source databases in a DDID-enabled format with Anonos-enabled CoTs holding JITI keys as well as information concerning which keys relate to which DDID data; Anonos-enabled CoTs may also contain data. Arrows represent data movement; data inputs and outputs contain different information.

Figure 6 below shows how Anonos BigPrivacy technology breaks the assumptions encoded in datasets which can be used to achieve re-identification by introducing entropy at the Data Subject (DS) level. This means that any given DS can map to any Data Attribute (DA), and any DA can map to any DS, all in the context that the dynamic de-identifiers for the DSs and DAs are just that: dynamic, so that the same underlying datum can have an unlimited number of different DDIDs representing itself, with respect to time, place, purpose, or any other criterion. Therefore, this is no longer a "1-to-k" mapping used by traditional privacy approaches, which vitiates data value, but a "k-to-k" mapping. This is entropy without the data value destruction that may constrain scientific and research advances based on data relationship discovery.

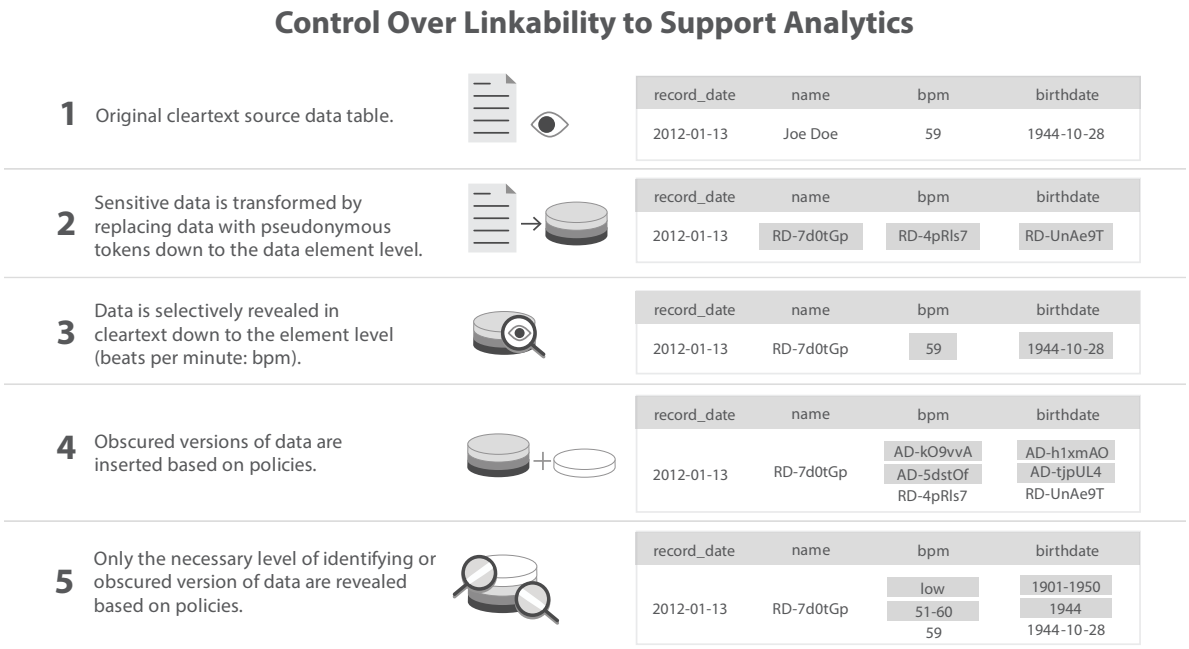


Figure 6

The ability to combine dereferencing identity with R-DDIDs and dereferencing associations with A-DDIDs to convey ranges or other associations/correlations in a non-identifying manner may decrease the need to distort, delete, or otherwise vitiate the data in data uses. This privacy-respectful data solution may therefore increase the accuracy of data analyses.[14]

14. See Anonos specific submissions by Sean Clouston, Ph.D., to the U.S. Federal Trade Commission at https://www.ftc.gov/policy/public-comments/2015/10/09/comment-00045

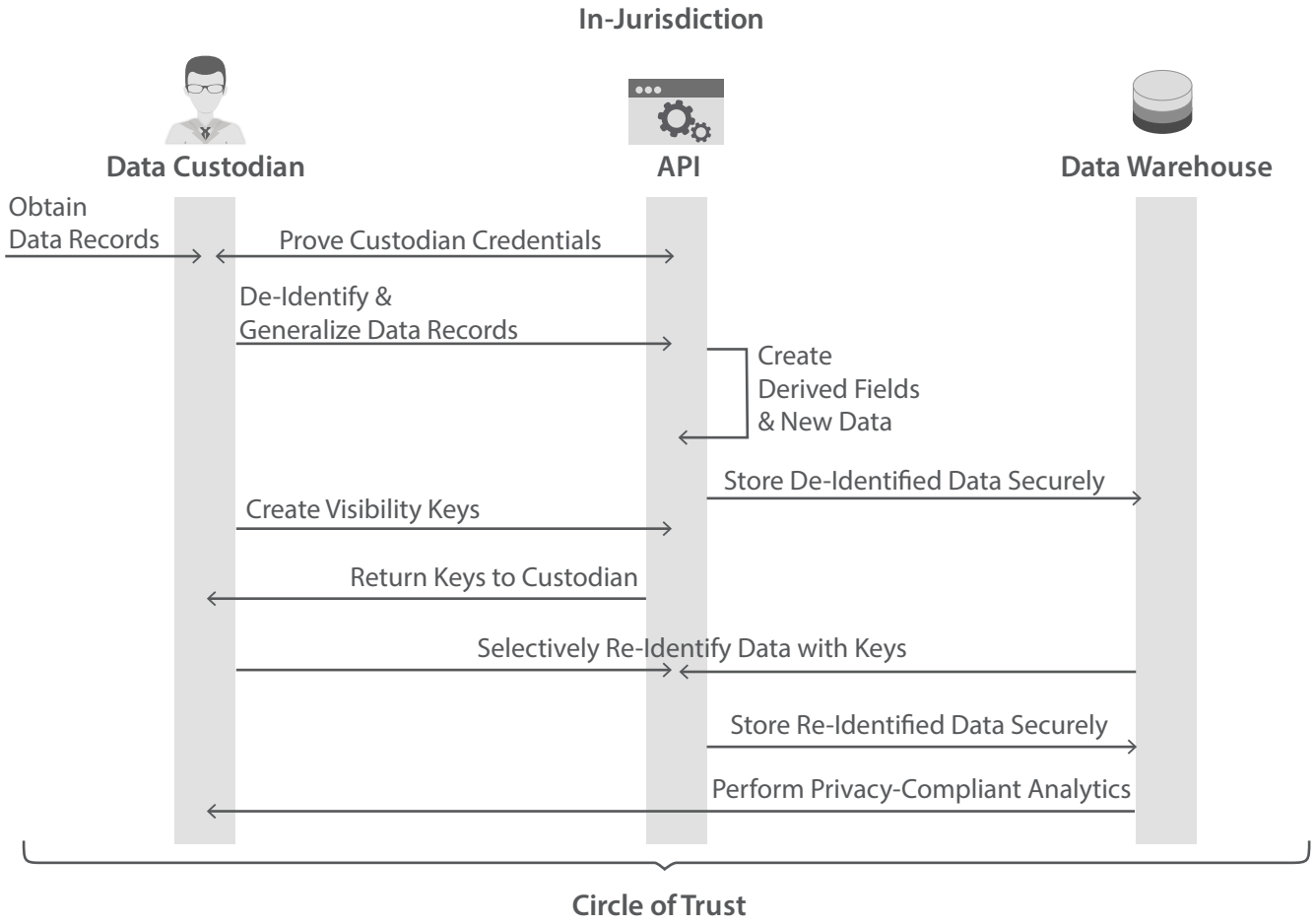
Appendix A: In-Jurisdiction Data Use and Analysis

A custodian's data must be protected to adhere to regulations of their local jurisdiction.

In the graphic below, data is "Anonosized" to comply with in-jurisdiction restrictions (e.g., data protection by default requirements in the EU).

In the EU, data protection by default requires privacy by design to be applied at the earliest opportunity (e.g., by pseudonymizing data) so that **by default** data access and use is limited to the **minimum extent and time necessary** to support specific authorized uses.

Anonos BigPrivacy technology embodies technical and organizational measures to enable statistical re-identification risk analysis supporting mathematical proof of low likelihood of re-identification.



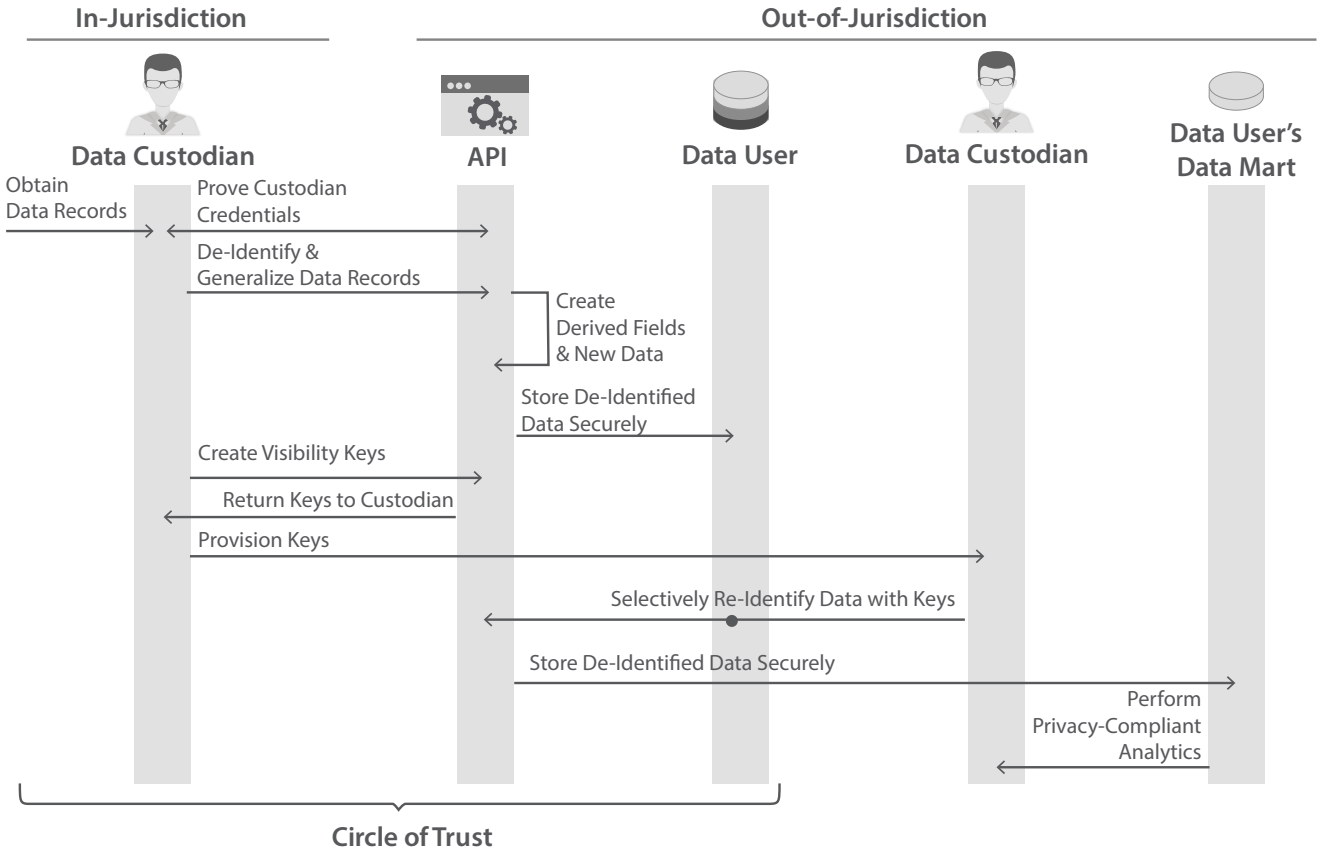
Appendix B: In- and Out-of-Jurisdiction Data Use and Analysis

A custodian's data must be protected (e.g., data protection by default in the EU) to adhere to regulations both internal and external to their local jurisdiction.

In the graphic below, data is "Anonosized" to comply with in-jurisdiction restrictions (e.g., data localization laws), data transfer restrictions, and out-of-jurisdiction data use restrictions.

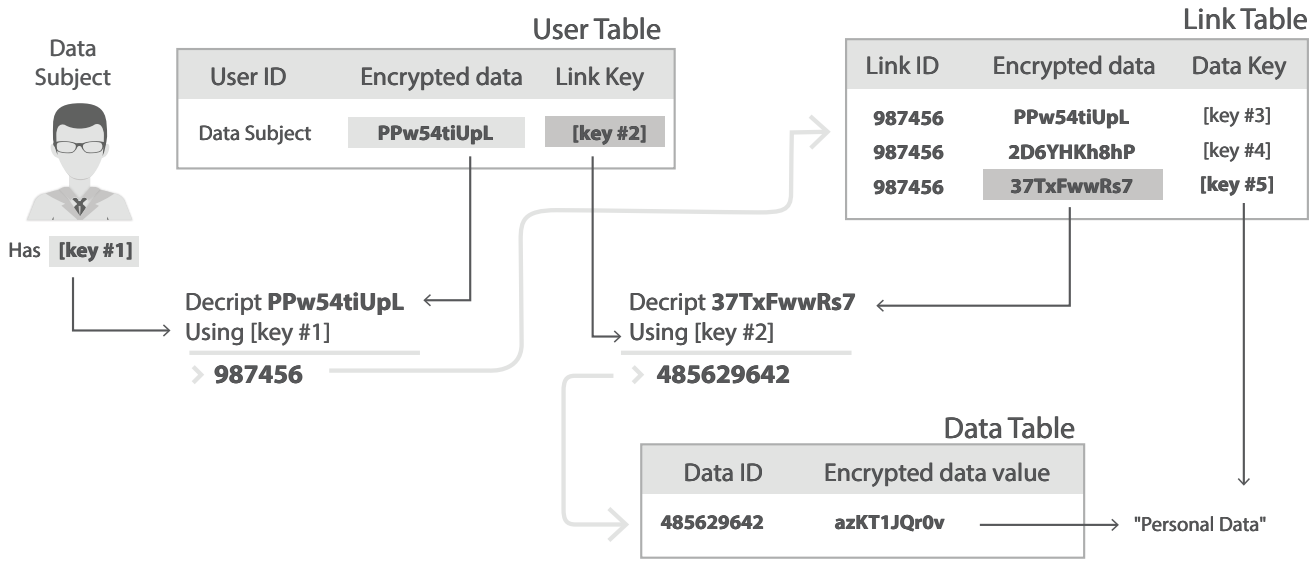
The custodian has in-jurisdiction access to any and all fields of data. Out-of-jurisdiction data user visibility is restricted to specific fields based on technically enforced privacy policies.

The data is protected against breach and privacy violations both inside and outside the governing jurisdiction.



Appendix C: Data Subject-Controlled De-Linking and Re-Linking of Data

The Anonos BigPrivacy platform can be used to implement de-linking and re-linking of dynamically-changing identifiers. By leveraging multiple layers of encryption, a Data Subject can control the sharing of data only with the Data Subject's explicit consent. The illustration below highlights how dynamically-changing identifiers may be stored in a non-identifying manner while providing Data Subjects with the ability to securely share specific personal data elements with another party.



In the illustration above, the Data Subject desires an improved experience at a website and, for that purpose, is willing to consent to the re-linking of some personal data by the website operator. The Data Subject provides unambiguous and explicit consent to the use of this data by giving the operator a key that may be used to re-link a specific set of the Data Subject's personal data. Afterward, the Data Subject may revoke the key manually at any time or automatically when a condition, such as the key expiring on a specific date or the passing of a specific period of inactivity, is met.

When the Data Subject provides consent (i.e., key #1), the following actions are performed by the BigPrivacy platform on the Data Subject's behalf to share the text value "Personal Data" (lower-right of the illustration) with the website operator:

1. The Data Subject's private key (key #1) is used to decrypt the Encrypted Link ID, yielding the Link ID of 987456 and the private key to be used in the Link Table (key #2). The Link ID and key #2 are embedded in a Privacy Rights Management (PRM)-enabled container that optionally enforces expiration dates and other security measures, and the container is sent to the website operator for future use. The operator is unable to link the contents of the container to the Data Subject because the Link IDs are random data bearing no relationship to the Data Subject.

2. Note that the Link Table contains three distinct records for Link ID 987456. Using key #2, which was provided to the website operator, only one of the Link Table records may be decrypted.

Successful decryption reveals the result 485629642 and another private key, key #5.

3. In the same manner as revealing the data in the Link Table (step #2), the Data ID 485629642 and key #5 are used to locate and decrypt the final data element in the Data Table that was requested by the website operator, revealing the text "Personal Data."

Benefits of Data Subject-Controlled De-Linking and Re-Linking of Data using BigPrivacy:

- Data links are explicit and always controlled by the Data Subject or a Trusted Party on behalf of the Data Subject.
- Data patterns are impossible to discern without private key access.
- Deleting or stopping the collection of valuable data or data links when regulations change is not needed since data linkage rules can change to enforce new requirements.
- Data can be collected securely prior to links being created.
- Data can be shared in a granular fashion, down to the individual data element level.
- Multi-layered encryption can be extended to any number of levels to create a granular permission and authorization hierarchy.
- Keys, including "Association Keys" (AKs) and "Replacement Keys" (RKs), as noted in Anonos U.S. Patent Nos. 9,361,481; 9,129,133; 9,087,216; and 9,087,215, are leveraged.
- Data is re-associated either "on-the-fly" or in time-limited databases using Keys including "Time Keys" (TKs) as noted in Anonos U.S. Patent Nos. 9,361,481; 9,129,133; 9,087,216; and 9,087,215.