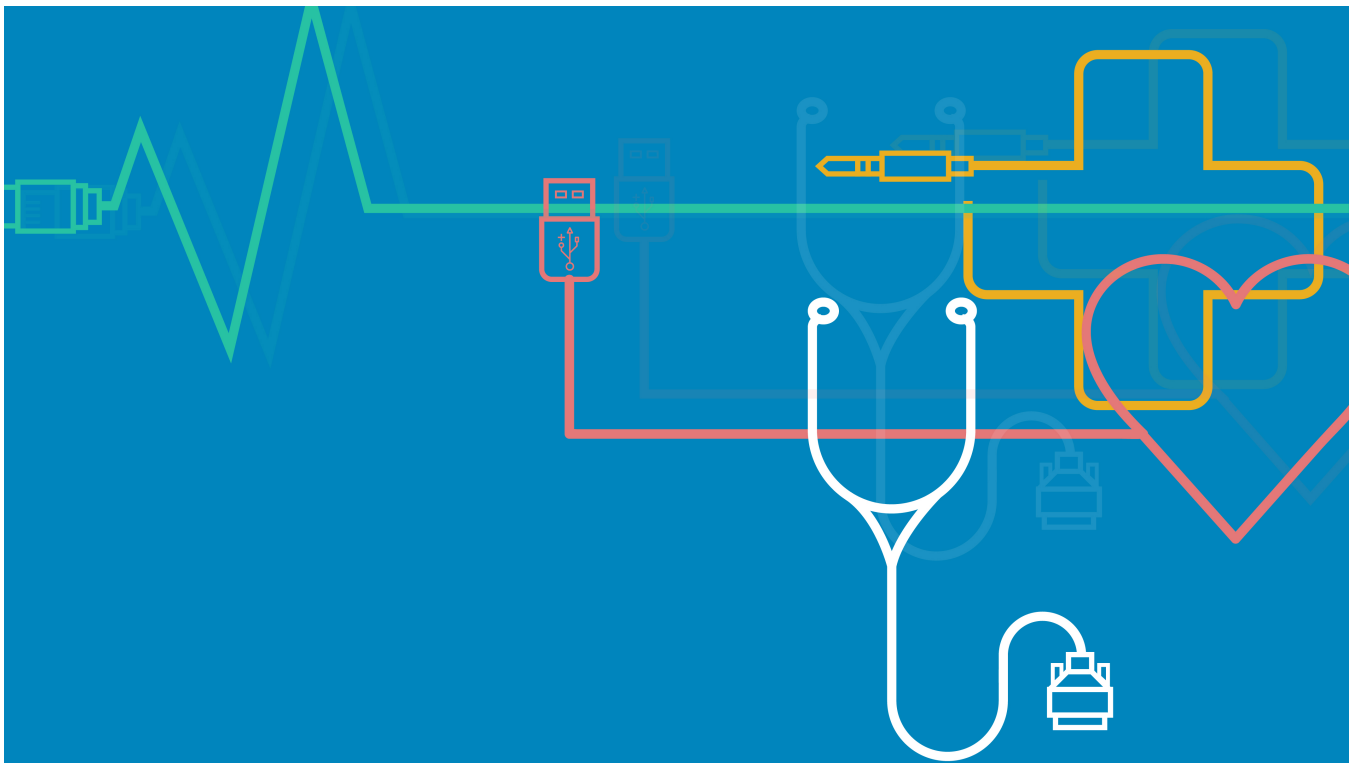


Big Data in Healthcare and Life Sciences

Anonos BigPrivacy Technology Briefing

Jonas Almeida, Ph.D.
Biomedical Informatics Department of Stony Brook University

Sean Clouston, Ph.D.
Department of Family, Population, and Preventive Medicine and Core Faculty in the Program in Public Health at Stony Brook University

Gary LaFever
Anonos Inc.

Ted Myerson
Anonos Inc.

Sandeep Pulim, MD
HealthXL and @Point of Care

April 2017 v2

Anonos®

Anonos® BigPrivacy® Technology

Unlocks the Value of Health Big Data by:

- **Maximizing Data Value:** Better control over sharing and use of restricted data (including health data) requires technology that supports emerging "data protection by default" requirements.

- **Minimizing Risk of Re-Identification:** Data analytics, artificial intelligence and machine learning ("Big Data") overpower the ability of traditional approaches to data security and privacy due to significant gaps in protection.

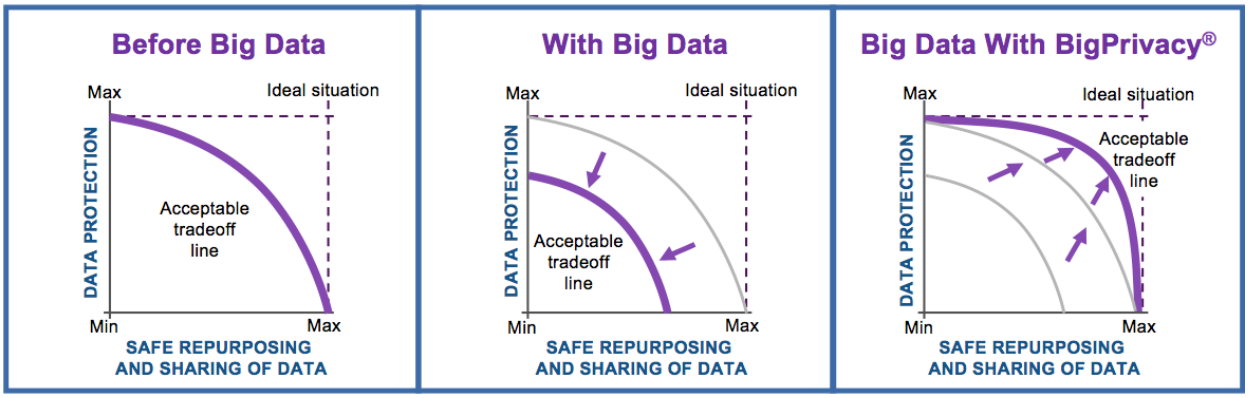
- **Enforcing Granular Controls:** privacy-respectful sharing and use of restricted data requires technically enforced granular controls that fuse together data security and privacy capabilities.

Anonos®

TABLE OF CONTENTS

1. Maximizing Value by De-Risking the Sharing and Repurposing of Restricted Big Data

In order to maximize the value of restricted[1] data, one must share and use that data for new purposes. Nonintegrated approaches to data security and privacy leave significant gaps in protection precisely because they serve very different purposes. Therefore, to effectively manage the liability risks from sharing and repurposing restricted data, one must fuse together data security and privacy.



Although greater use of data increases its value, this graph shows that even before widespread use of data analytics, artificial intelligence and machine learning ("Big Data"), concerns about the effectiveness of nonintegrated approaches to security and privacy for restricted data ***limited the safe repurposing and sharing of that data.***

The acceptable tradeoff line shows that the safe repurposing and sharing of restricted data was ***limited*** even before Big Data.

Safe repurposing and sharing of restricted data actually ***decreases due to compounding effects of:***

- Increasing volume, variety and velocity of Big Data; and
- Increasing risk of re-identification due to the Mosaic Effect.

Threats from compounding effects of increasing volume, variety and velocity of data and risks of re-identification due to the Mosaic Effect ***shrink*** the acceptable tradeoff line.

Integrated and simultaneous support for both security and privacy ***improves the overall effectiveness*** of data protection for restricted data thereby maximizing the safe repurposing and sharing of that data.

BigPrivacy granularized data protection ***expands*** the acceptable tradeoff line. This enables the sharing and repurposing of restricted data under controlled conditions to improve both privacy and security.

Security Shortcomings

Security is about *data access*: does it keep you in or lock you out? The dividing line is called the *perimeter*, so the focus is on perimeter controls, i.e., granting or denying access to entire data stores. Whether the perimeter is established at a physical structure level, at the machine or device level, or even at the application level, security is an all-or-nothing proposition: you are either outside the perimeter or inside it. But once authorized people are allowed access inside the perimeter, security controls do not limit what they can do with that data, because the perimeter is only a first line of defense. Since security technologies were not designed to support fine-grained, *granularized* control down to the data element level, something else is needed: privacy.

[1] Restricted data is data subject to legal limitations on processing, sharing, storage and/or other uses. Examples of restricted data include "protected health information" or "PHI" under the U.S. Health Insurance Portability and Accountability Act (HIPAA) and "personal data" under the EU General Data Protection Regulation (GDPR).

Privacy Shortcomings

While security is about controlling access, privacy is about using policies and contracts to control data use by already authorized parties. However, when the data is shared and repurposed at larger scales, these policies and contracts become ineffective. While there are traditional approaches to privacy that do not rely exclusively on policies and contracts (e.g., de-identification and differential privacy), these traditional approaches degrade the data's quality because they are designed to introduce "noise" to prevent precision – *or identifying* – use of that data. They further fail to address large scale data sharing and repurposing because seemingly "anonymous" data sets can often be combined and analyzed to reveal restricted or sensitive information. This dangerous risk is called the *Mosaic Effect*.

Anonos BigPrivacy Integrates Security and Privacy

Anonos BigPrivacy technology encapsulates data to integrate security and privacy to maximize value from sharing and repurposing of restricted data to enable granular, contextual, and programmatic control over data. BigPrivacy minimizes liability risks from sharing and repurposing restricted data by supporting granularized data rights management. This process, called "Privacy Rights Management" or "PRM," de-risks the data.[2] BigPrivacy de-risks data by enabling the selective locking and unlocking of data at any level – all the way down to the data element level. Anonos BigPrivacy technology thus:

- ***Complements Security*** – If other security techniques fail, the exposed data does not have any value or meaning by itself, so it cannot be used to re-identify the subject. Therefore, since there is nothing to be gained by trying to obtain this useless data, the data itself is at less risk.
    - In fact, the only way to determine the value or meaning of data that has been de-risked using BigPrivacy is under controlled conditions via access to security keys, without which the data has no meaning or intelligibility.
- ***Complements Privacy*** – It does not nullify the need for other approaches to privacy, such as:
    - Policies and contracts, which *BigPrivacy technologically enforces*.
    - Traditional Privacy Enhancing Techniques (PETs), *BigPrivacy reduces the level of errors introduced and helps maintain the very low desired levels of re-identification risk – even when "anonymous" data sets are combined.*

[2] Ted Myerson, Co-Founder of Anonos, presented a **TED** Talk on the healthcare and life science benefits of BigPrivacy technology enforced Privacy Rights Management. A video of, and the transcript for, this TED Talk is available at https://anonos.com/TEDTalk. **TED** Talks is a trademark of **TED** Conferences, LLC.

Anonos BigPrivacy technology de-risks the sharing and repurposing of restricted data. What drives the need for this de-risking is that the liability risk from data misuse, abuse or compromise increases exponentially when restricted data is shared or repurposed at scale. And since data is shared and repurposed at scale across its entire lifecycle, BigPrivacy acts at every point in and across that lifecycle – without compromising privacy – by technologically:

1. Restricting data use to granularized authorized purposes by authorized parties: this maximizes retained value;

2. Enabling broader use of data by imparting value, all without revealing unnecessary identifying information; and

3. Limiting re-identification capabilities to authorized parties only.

Unlocking the Value of Health Data

Anonos BigPrivacy technology retains the full value and utility of restricted data to support authorized use cases, all while minimizing the risk of data misuse, abuse or compromise. We call this process "Anonosizing" data.

2. Overview of BigPrivacy for Healthcare/Life Sciences

The burgeoning volume and variety of data available in healthcare and life sciences translates into both unparalleled opportunities for deriving value and unparalleled risks for data security and privacy.

The Cancer Moonshot, Precision Medicine Initiative, and other potential breakthrough efforts all depend on access to and use of restricted health data. However, nonintegrated data security and privacy methodologies lack the capabilities to enable this access and use.

A new approach to data protection is required, one that integrates data security and privacy by dynamically encapsulating data elements until they are needed. Highly granularized data elements can be kept safely protected by using dynamically changing pseudonymous identifiers, making it impossible to discover data values until they are revealed under controlled conditions. Anonos is the first company to develop such data protection technology, which it markets as "BigPrivacy."

Through BigPrivacy, healthcare and life sciences can derive the maximum value from data without compromising the data's security, privacy, accuracy, or utility.

Two fundamental factors stand in the way of realizing potentially breakthrough healthcare/life science initiatives (e.g., the Precision Medicine Initiative, Cancer Moonshot, etc.):

1. The fact that traditional data protection (i.e., security and privacy) technologies were invented before the era of Big Data – data volumes are exploding, more data has been created in the past few years than in the entire previous history of the human race; by 2020, at least a third of all data will pass through the cloud (a network of servers connected over the Internet)[3]; and
2. The inability of nonintegrated data protection technologies to resolve a phenomenon that mathematicians and statisticians call the "Mosaic Effect" – i.e., the more data sources that exist, the easier it becomes to unearth an individual

[3] https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#e6f819317b1e

data subject's identity *without having any access to the data subject's primary identifiers (e.g., name, data of birth, street address, etc.)*.

The failure to overcome these shortcomings:

- Increases vulnerability to ***security breaches***;
- Risks the ***loss of patient control*** over sensitive health data;
- Exposes sensitive health information to ***misuse and abuse***;
- ***Threatens confidential communications*** between doctors and patients; and
- ***Endangers access to accurate information*** necessary for new insights and discoveries.

Significant domestic and international attention has been focused on attempts to use the principles of De-Identification[4] and Data Protection by Default[5] for reconciling conflicts between data use/utility and data security/privacy. The following are examples of such initiatives:

- **EU GDPR Data Protection by Default:** – Protection at the earliest opportunity;[6]
- **Article 29 Working Party EU-U.S. Privacy Bridges Project:** Bridge 7 - De-identification of personal data to help bridge differences between EU and U.S. data protection regimes;[7]
- **Privacy Commissioner of Canada** (Daniel Therrien): Consideration of de-identification as an alternative to consent under the Personal Information Protection and Electronic Documents Act; [8]

[4] De-identification is a process of removing or obscuring personal information from data in an attempt to resolve conflicts between data use/utility and data security/privacy. In 2015, the Information Access Division of the National Institute of Standards and Technology (NIST) published a report on de-identification (see http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf - "NIST De-Id Report"). In June 2016, the Information and Privacy Commissioner of Ontario ("IPC") published de-identification guidelines for structured data (see https://www.ipc.on.ca/images/Resources/Deidentification-Guidelines-for-Structured-Data.pdf - "IPC De-Id Report").
[5] Privacy by Design ("PbD") is the approach developed by Ann Cavoukian, Ph.D., former IPC Commissioner for embedding privacy into the system design process (see https://www.ipc.on.ca/english/Privacy/Introduction-to-PbD/). Data Protection by Default, the most recent implementation of PbD, is required under Recital 78 and Article 25 the EU General Data Protection Regulation ("GDPR"); it requires PbD techniques to 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 a specific product or service authorized by an individual data subject.
[6] See *id*.
[7] https://privacybridges.mit.edu/
[8] https://www.priv.gc.ca/information/research-recherche/2016/consent_201605_e.asp#heading-0-0-6-2

- **U.S. Health and Human Services (HHS):** Protect health information by improving trust in de-Identification methodologies and reducing the risk of re-identification;[9]
- **Future of Privacy Foundation ("FPF"):** Proposed schema with ten gradations of de-identification;[10] and
- **Information Accountability Foundation ("IAF"):** Leveraging Dynamic Data Obscurity (a term for selectively enforcing acceptable use policies) to isolate data until it is deemed appropriate for fair processing.[11]

Anonos® BigPrivacy® technology is a both ***a foundational new approach*** and ***a significantly improved solution*** that integrates and simultaneously supports both security and privacy to enable granularized data protection.

Anonos BigPrivacy Value Statement:

Data sharing and repurposing is at the core of improving health analytics both inside and outside the clinic, but reasonable and increasing concerns about gaps left by nonintegrated security and privacy increase the difficulty of this effort.

BigPrivacy technology responds to these needs by integrating and simultaneously supporting both security and privacy, thus enabling data protection that dynamically flows with data. The result is the technological enforcement of policies over the full lifecycle of data.

BigPrivacy technology makes more data available while simultaneously enhancing data security and privacy to enable compliance with ranges of policies for protecting individuals and their data.

By granularizing data controls so that re-identification risk is minimized, Anonos BigPrivacy technology enables entities to:

De-Risk Data. Discover Value.®

[9] Section 6.3 of https://www.healthit.gov/sites/faca/files/HITPC_Health_Big_Data_Report_FINAL.pdf - "Health Big Data Recommendations Report."
[10] https://fpf.org/2016/04/25/a-visual-guide-to-practical-data-de-identification/
[11] http://informationaccountability.org/category/dynamic-data-obscurity/

BigPrivacy technology enables data protection policies to dynamically flow with data so that disparate requirements of sovereign national (e.g., U.S. and UK) and international (e.g., EU) entities can be enforced at the required levels of necessity and proportionality. By "de-risking" data at the earliest opportunity, BigPrivacy enables data to flow within, between and among applications and platforms in a manner that embodies Data Protection by Default and Pseudonymisation (such as required under the GDPR). This improves both data security and privacy by supporting granularized data protection. Anonos BigPrivacy provides technical protection over the full lifecycle of data by supporting pseudonymisation at the earliest point possible, thereby supporting greater trust and confidence both domestically and internationally.

Over the full life cycle of data, Anonos BigPrivacy technology:

- **Maximizes** authorized uses of data while minimizing unauthorized uses of data by minimizing re-identification risks;
- **Facilitates** compliance with and auditability against 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 entities, industries, states, countries, regions, etc. – 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.

BigPrivacy technology has applications in numerous vertical industries and geopolitical jurisdictions. It holds particular promise in the context of U.S. healthcare because of the sensitive nature of health data and the ***inability of traditional nonintegrated data security and privacy to comply with requirements under the HIPAA[12] Security Rule and Privacy Rule, a problem due to the increasing volume, velocity and variety of data.***

[12] U.S. Health Insurance Portability and Accountability Act of 1996, as amended.

This Briefing focuses on the following uses of BigPrivacy technology specific to U.S. healthcare:

1. **Health Information Exchanges**
2. **Minimum Necessary Requirements**
3. **Precision Medicine**
4. **Data Breaches**
5. **Unstructured Data/Cohort-Based Research**
6. **Blue Button Initiative/Patient Reported Outcomes**
7. **Revenue Cycle Management**
8. **Fraud and Abuse Mitigation**
9. **Data Minimization**
10. **Cancer Moonshot/Genetic Research**

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 he also suggested that Information Age healthcare technologies could be predictive, preemptive, precise, and participative.[13] To support Information Age healthcare, Anonos believes that granularized data protection must be embedded at the data element level in order to reduce the risk of re-identification from the increasing volume, variety and velocity of data in today's data-driven society.

BigPrivacy's patented systems and methods for granularized data protection overcome the inabilities of traditional nonintegrated security and privacy, thus defeating the Mosaic Effect and supporting the maximum extraction of value from data.

[13] Hesse, B. W., Ahern, D. & Beckjord, E. (2016). *Oncology Informatics: Using Health Information Technology to Improve Processes and Outcomes in Cancer*.

3. Tensions Between FIPPS and Big Data in Healthcare/Life Sciences

To understand the state of data protection today, one must start with the Fair Information Practice Principles ("FIPPs").[14] Rooted in the United States Department of Health, Education and Welfare's seminal 1973 report, *Records, Computers and the Rights of Citizens*,[15] the FIPPs are fundamental to many U.S. (Federal and state) and foreign (national and international) data protection regimes. The specific FIPPs are:

- **Transparency:** Organizations should be transparent and notify individuals regarding collection, use, dissemination, and maintenance of personally identifiable information ("PII").
- **Individual Participation:** Organizations should involve the individual in the use of PII and, to the extent practicable, seek individual consent for the collection, use, dissemination, and maintenance of PII. Organizations should also provide mechanisms for appropriate access, correction, and redress regarding use of PII.
- **Purpose Specification:** Organizations should specifically articulate the authority that permits the collection of PII and specifically articulate the purpose or purposes for which the PII is intended to be used.
- **Data Minimization:** Organizations should only collect PII that is directly relevant and necessary to accomplish the specified purpose(s) and only retain PII for as long as is necessary to fulfill the specified purpose(s).
- **Use Limitation:** Organizations should use PII solely for the purpose(s) specified in the notice. Sharing PII should be for a purpose compatible with the purpose for which the PII was collected.
- **Data Quality and Integrity:** Organizations should, to the extent practicable, ensure that PII is accurate, relevant, timely, and complete.

[14] http://www.nist.gov/nstic/NSTIC-FIPPs.pdf
[15] https://www.justice.gov/opcl/docs/rec-com-rights.pdf

- **Security:** Organizations should protect PII (in all media) through appropriate security safeguards against risks such as loss, unauthorized access or use, destruction, modification, or unintended or inappropriate disclosure.
- **Accountability and Auditing**: Organizations should be accountable for complying with these principles, providing training to all employees and contractors who use PII, and auditing the actual use of PII to demonstrate compliance with these principles and all applicable privacy protection requirements.

The White House report, *Big Data: Seizing Opportunities, Preserving Values*,[16] published in May 2014, highlights pressure exerted by Big Data on the FIPPs. A report published on the same day by the President's Council of Advisors for Science & Technology, *Big Data and Privacy: A Technological Perspective,* further emphasizes this very point.[17] In a follow-up to these two reports, the Privacy and Security Workgroup ("PSWG") of the Health Information Technology Policy Committee ("HITPC") was charged with investigating privacy and security issues related to the electronic exchange of health information and providing recommendations on their findings to the National Coordinator for Health Information Technology at the U.S. Department of Health and Human Services ("HHS"). In August 2015, the Health IT Policy Committee Privacy and Security Workgroup published the *Health Big Data Recommendations Report*[18] which reported on increasing Big Data concerns regarding detrimental global effects on personal privacy, particularly on health-related matters. This report underscores tensions inherent in trade-offs between data use/utility and data security/privacy, trade-offs harmful to the advancement of medical knowledge and positive health outcomes both domestically and internationally. Selected excerpts from the *Health Big Data Recommendations Report* follow:

> *Big data is blurring the lines between traditional health information (e.g., clinical or billing information) and other information (e.g., user-generated*

[16] https://www.whitehouse.gov/sites/default/files/docs/big data privacy report may 1 2014.pdf
[17] https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf
[18] See *supra* note 7.

information about diet, steps, workouts, sleep, and mood). Consequently, ***defining health information is becoming more difficult because almost all information has potential to, in some way, become health-related information, depending on how it is used.*** *Growth in the amount and availability of such information places additional pressure on core FIPPs.*[19]

Big data analytics and research begins with researchers examining trends and patterns in large data sets without first formulating a hypothesis. As a result, the need to gather as much information as possible before identifying a research purpose conflicts with longstanding FIPPs that require defining the specific purpose(s) for which information is collected and limiting the amount of personal information to what is necessary to accomplish the specified purpose(s). Regardless of the challenge posed by big data, panelists and PSWG members agreed that organizations should examine their collection and retention practices and be mindful of over collection.[20]

Big data introduces new risks of re-identification due to the volume of data and the broad variety of data sources in the big data ecosystem.

Some say the FIPPs are unsuited for the era of big data (e.g., analytical methods are putting pressure on traditional principles such as confidentiality, security, individual participation through meaningful patient consent, transparency and data minimization (including collection, use, and purpose limitation).

Nevertheless, presenters defended the FIPPs, stating they still provide "a strong, standardized structure that promotes responsible and efficient use of data while allowing for innovations in analytics and application."

The security threat landscape changes constantly over time. These evolving security threats are driven by vulnerabilities that arise from designing and

[19] *id* at Section 4.1.
[20] *id* at Section 4.1.5.

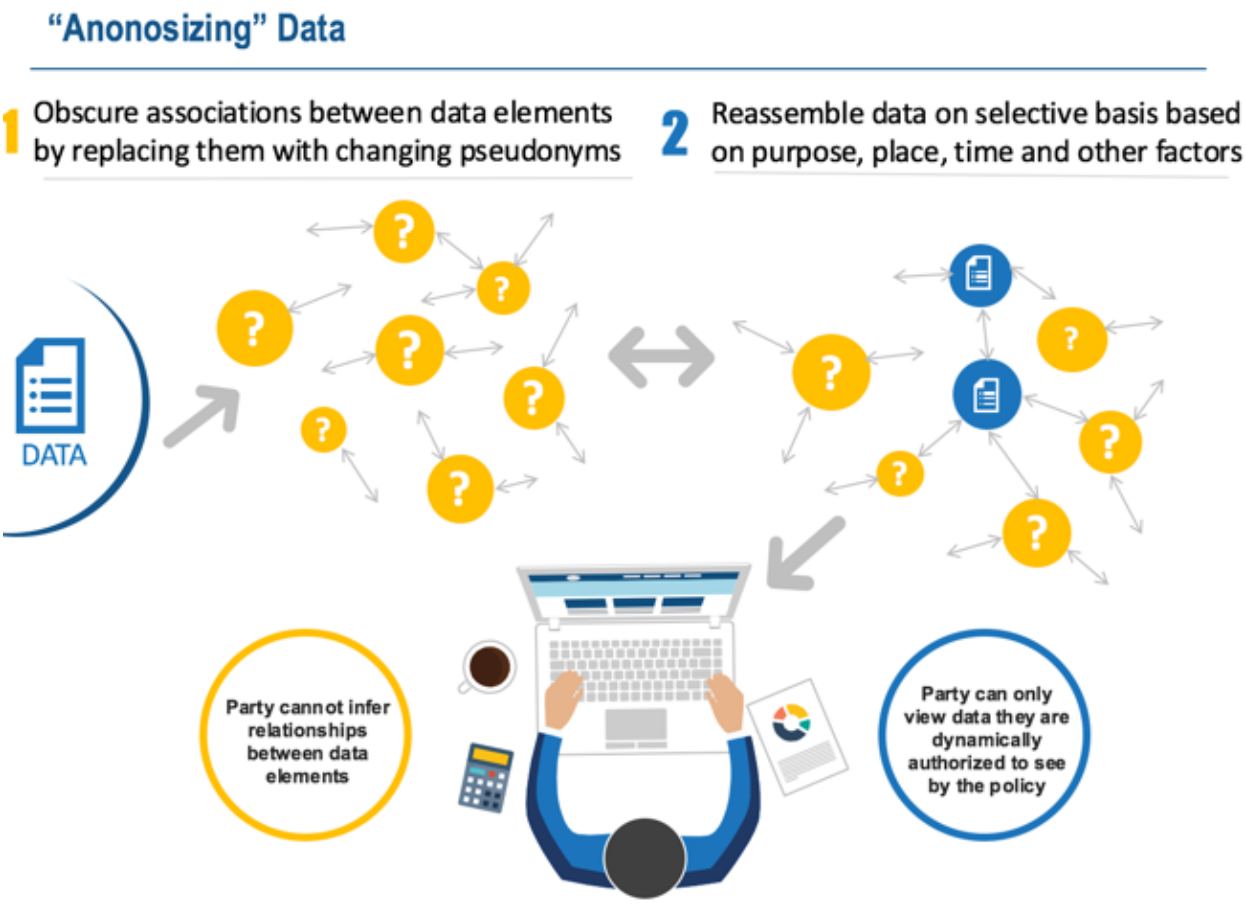
deploying highly complex software and hardware. ***Ultimately, there is no such thing as zero risk. In response to this complexity, organizations should adopt a balanced, holistic approach to security that looks at operations end-to-end and applies a risk-based framework.*** *This holistic approach should include considerations like physical security.*

HIPAA defines high-level objectives, but panelists stated the need for a risk-based framework that defines very specific, contextual, and evolving controls that are applied to reduce risk to an acceptable level. "The only pragmatic way to secure data in healthcare and in any other domain is to consistently follow an industry developed risk-based framework."[21] (emphasis added)

[21] *id* at Appendix B.

4. How BigPrivacy De-Risks Data to Discover Value

Anonos BigPrivacy technology “de-risks” data (i.e., severely minimizes the likelihood of re-identification, without undermining the research or other value of the underlying data) by supporting a risk-based framework that technologically and programmatically enforces data protection (i.e., data security and privacy) policies in a contextually flexible, selective manner. BigPrivacy achieves this all the way down to hierarchically lower data element levels and even down to the individual data element level.



Over the full lifecycle of data, BigPrivacy technology:

- Maximizes authorized uses of data while minimizing unauthorized uses of data – all by minimizing re-identification risks;
- Facilitates compliance with and auditability against 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 companies, industries, states, countries, regions, etc. – 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.

Selective Access to Metadata and Controlled Linkability of Data

The following diagram shows the difference between the status quo and BigPrivacy. In Figure 1, the spheres on the left represent data elements grouped according to metadata parameters (cf., metadata is data that provides information about other data). Here, the metadata reveals interrelationships between and among the top three spheres and between and among the bottom four spheres (where each sphere represents a data element). These interrelationships enable tracking, profiling, inferences, deductions, analyses, understanding and correlative relationships represented by the dotted lines between and among the spheres on the left side of the figure.

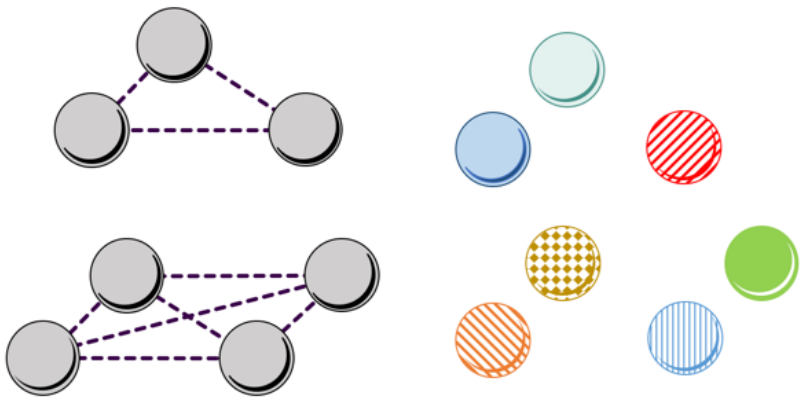


Figure 1

In contrast, on the right side of Figure 1, each of the different designs on the spheres represents a unique Dynamic De-identifier® ("DDID®") (as more fully described below) that has been used to replace the data element represented by the sphere. Using different DDIDs means that little or no metadata exists or relates to any of the spheres on the right side of the figure. While interrelationships between or among the right-side spheres may exist, there is no way to identify or infer them, because the relevant information has been removed or made undiscoverable. Only through access to BigPrivacy Just-In-Time-Identity® (or JITI®) keys/schemata ("JITI keys") necessary for transforming DDIDs into an intelligible form, an access which is highly restricted and controllable, can this information be made available. Consequently, the replacement of data elements with DDIDs significantly increases the difficulty of attempting to track, profile, infer, deduce, analyze, understand or establish correlations between or among any of the spheres representing data elements without authorization to do so.

Programmatic Enforcement and Auditability

Granular, contextual, programmatic enforcement by BigPrivacy on the front-end makes it easier to audit compliance with data protection policies on the back-end. This increases the accountability and trust necessary for a wide-scale, domestic and international acceptance of data analysis and use that simultaneously maximizes the value of data and ensures both protection for that same data and respect for the rights of individual data subjects. The same data may be subject to different jurisdictional requirements based on the source and/or use of the data. For example, depending on how the data is captured, data representing a heart rate reading (e.g., 55 beats per minute) may be subject to different data protection policies. If the data is captured by means of a personal health device in the U.S., use of the data may be subject only to terms and conditions of the device and/or application software used to capture the information. If, however, the data is captured in connection with providing healthcare services in the U.S., use of the data may be subject to federal HIPAA and applicable state laws. In a third case, if the data is captured in connection with federally funded

research in the U.S., use of the data may be subject to the Common Rule.[22] As a result, scalable programmatic data protection technology solutions, such as BigPrivacy, are needed for multiple reasons, including accommodating jurisdictionally disparate data protection policies of different business, industry, government, regulator and/or other stakeholder group(s).

Benefits of BigPrivacy Over Traditional Data Protection Techniques

In contrast to the BigPrivacy approach where granularized controls enable the maximization of data protection and data value, traditional approaches to data protection are generally binary: either data protection is maximized at the sacrifice of data value; or data value is maximized at the sacrifice of data protection. The BigPrivacy approach exposes this as a false dichotomy. For example, efforts to improve data security by encrypting data result in data being protected but unusable in its protected form; or, conversely, in the data's becoming vulnerable when it is decrypted for the very purpose of enabling use. Figure 2 compares the impact of traditional approaches to data protection on the preservation of data value versus the preservation (or expansion) of data value using BigPrivacy.

[22] http://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html

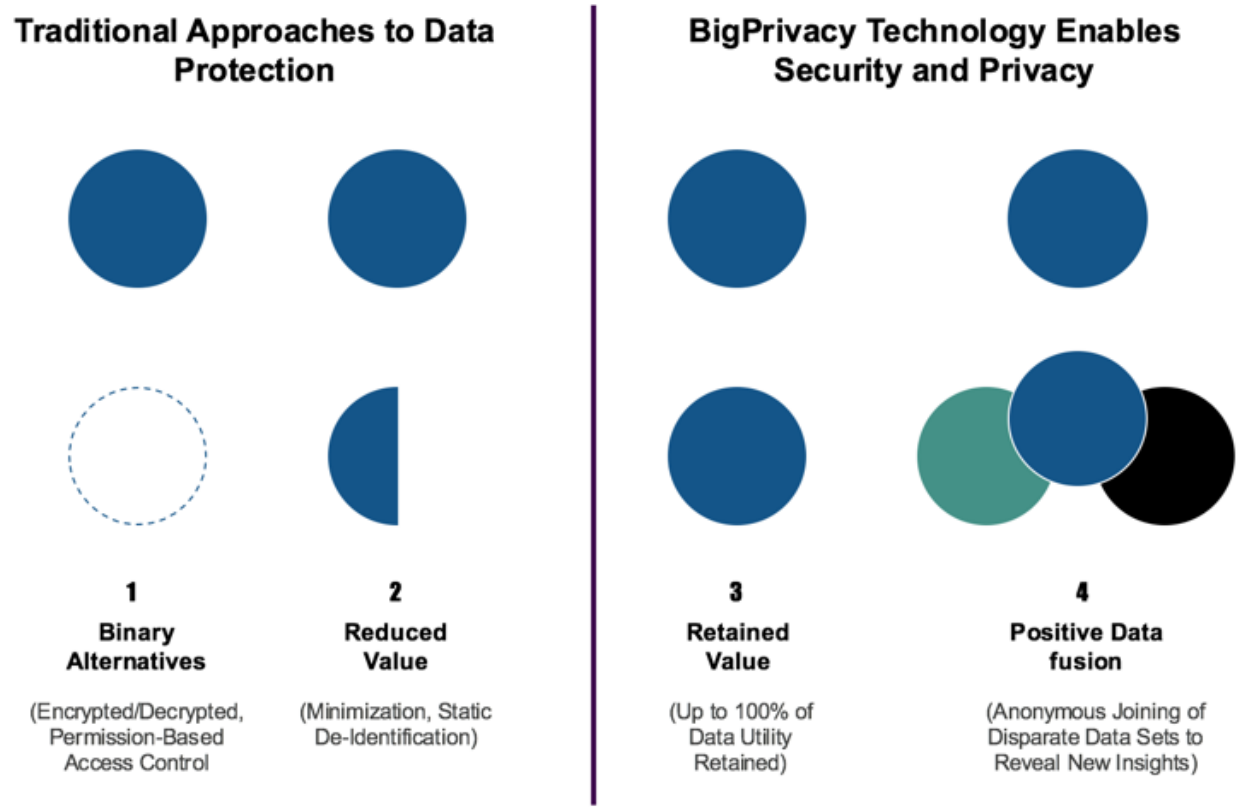


Figure 2

Column #1 in Figure 2 represents the effect of binary alternatives (e.g., encryption) wherein the top blue sphere shows the value of original data (in unprotected form) and the dotted sphere represents the data value when that data is in a protected form, rendering it unusable (the empty sphere shows that the data value is essentially zero).

Column #2 in Figure 2 illustrates the reduction in data value due to (i) removing data from the ecosystem in response to concerns over using data for purposes other than those primarily intended ("Data Minimization") and (ii) from using traditional static (and data value-reducing) approaches to obfuscating data in order to achieve de-identification.

Column #3 in Figure 2 shows that BigPrivacy retains up to 100% of data value.

Column #4 in Figure 2 illustrates that using BigPrivacy enables the possibility of positive data fusion by enabling data protection to dynamically flow along with the data to technologically enforce polices *over the full lifecycle of data*.

Other approaches to simultaneously supporting data security and privacy require the continual use of a proprietary "lens" (e.g., browser, application or platform). BigPrivacy is different, however, because it does not require this at all. Instead, by enabling data protection to dynamically flow along with data, BigPrivacy ensures that the protected form of the data can flow within, between and among all applications and platforms.

BigPrivacy also delivers immediate benefits to existing business and technology practices – without modifying those practices. By using DDIDs, current systems and processes (e.g., data analytic engines) are intentionally rendered unable to recognize relationships between and among dissociated and/or replaced data elements. The result is that data analytic engines and the like can process information using existing capabilities – but without creating inferences, establishing correlations, instantiating profiles or deriving conclusions – except to the extent they are expressly authorized to do so (using JITI keys) by data subjects or by authorized third parties ("Trusted Parties").

Analysis of BigPrivacy De-Risking Capabilities

Figure 3 represents two phases for de-risking data with BigPrivacy. The first phase (above the horizontal dividing line) highlights the elimination of visible links between data elements so a party cannot infer or deduce relationships between data elements. Rendering data elements as DDIDs dynamically obscures cleartext source data and the resulting technologically enforced obscurity flows along with the data within, between and among applications and platforms. Data rendered with DDIDs is still present but, from an information theoretic perspective, the knowledge or context necessary to understand the data becomes dissociated from the data by means of JITI keys – i.e., the DDIDs need not contain any information about the underlying data element(s). In

the second phase (below the horizontal dividing line), JITI keys are assigned to allow selective disclosure of data based on JITI key-enabled policy controls (e.g., purpose, place, time and/or other designated trigger factors); in selectively revealing data, the level of detail/clarity provided to each key holder – e.g., original cleartext, perturbed value, summary information, etc. – can also be dynamically controlled. Notably, there are no limits on the number of different selective disclosures that can be made serially or in parallel; on the number of different authorized users to which any one or more of the disclosures can be made; or on the constraints or policies (such as time, purpose, place, other (association, relationship, quantitative), etc.) governing such disclosures.

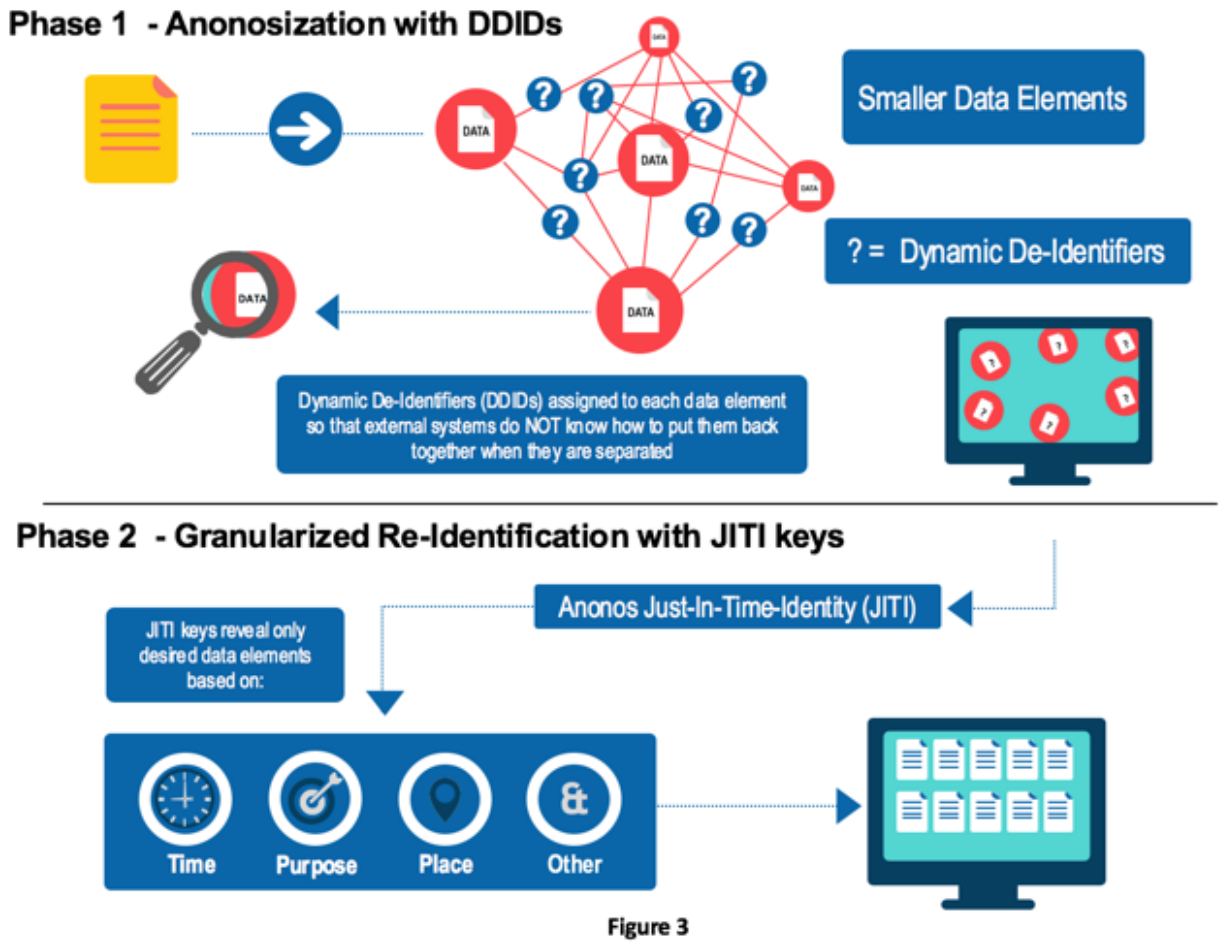


Figure 3

Granular, contextual, programmatic enforcement of data protection policies with BigPrivacy supports the statistical assessment of (i) the probability that a data breach and/or data re-identification will occur or (ii) the rank ordering of such incidents (i.e.,

non-parametric methods). BigPrivacy is more efficient from an information theoretic perspective than other approaches to protecting data because the value of the data is still accessible but the underlying identifying information is not. In other words, the identifying information has no leakage, meaning zero identifying information is leaked, while the value of the data is safely and intentionally "leaked," in a positive, permitted way (which may itself be subjected to standard information theoretic optimizations), meaning the value is made available only to those who are authorized to use it and that even such limited use may be constrained further by time, person, place, etc.

The granular, contextual, programmatic structure of BigPrivacy technology significantly reduces the probability of a data breach or re-identification; this can be mathematically proven. For an example of such a proof of BigPrivacy's effectiveness in de-risking data, see Anonos-specific submissions by Sean Clouston, Ph.D., to the U.S. Federal Trade Commission,[23] which concluded that data replaced with DDIDs down to the data element level (a process we refer to as "Anonosizing®" data) results in no greater probability of re-identification than guessing the identity of highly encrypted data.

However, unlike encrypted or other non-Anonosized data which vitiate or eliminate the value implicit in data, Anonosized data in its protected form can still be used to generate maximal value. In addition: (a) different DDIDs can be assigned to the same data element(s) at different times and/or different places and/or different purposes and/or according to other criteria, thus making it exceedingly difficult (near-zero probability from a mathematical perspective) for parties not in possession of JITI keys to track, profile, infer, deduce, analyze or otherwise understand protected data; and (b) the same DDID(s), if expired for any reason, can be (but are never required to be) assigned to different data elements, also at different times and/or different places and/or different purposes and/or according to other criteria, thus making it equally difficult for interlopers or other "bad actors" to establish any meaningful continuity or audit trail, since these reassigned DDIDs would then refer to data elements that bore no meaningful

[23] https://www.ftc.gov/policy/public-comments/2015/10/09/comment-00045

relationship, correlative or otherwise, to any and all data elements to which they had been assigned.

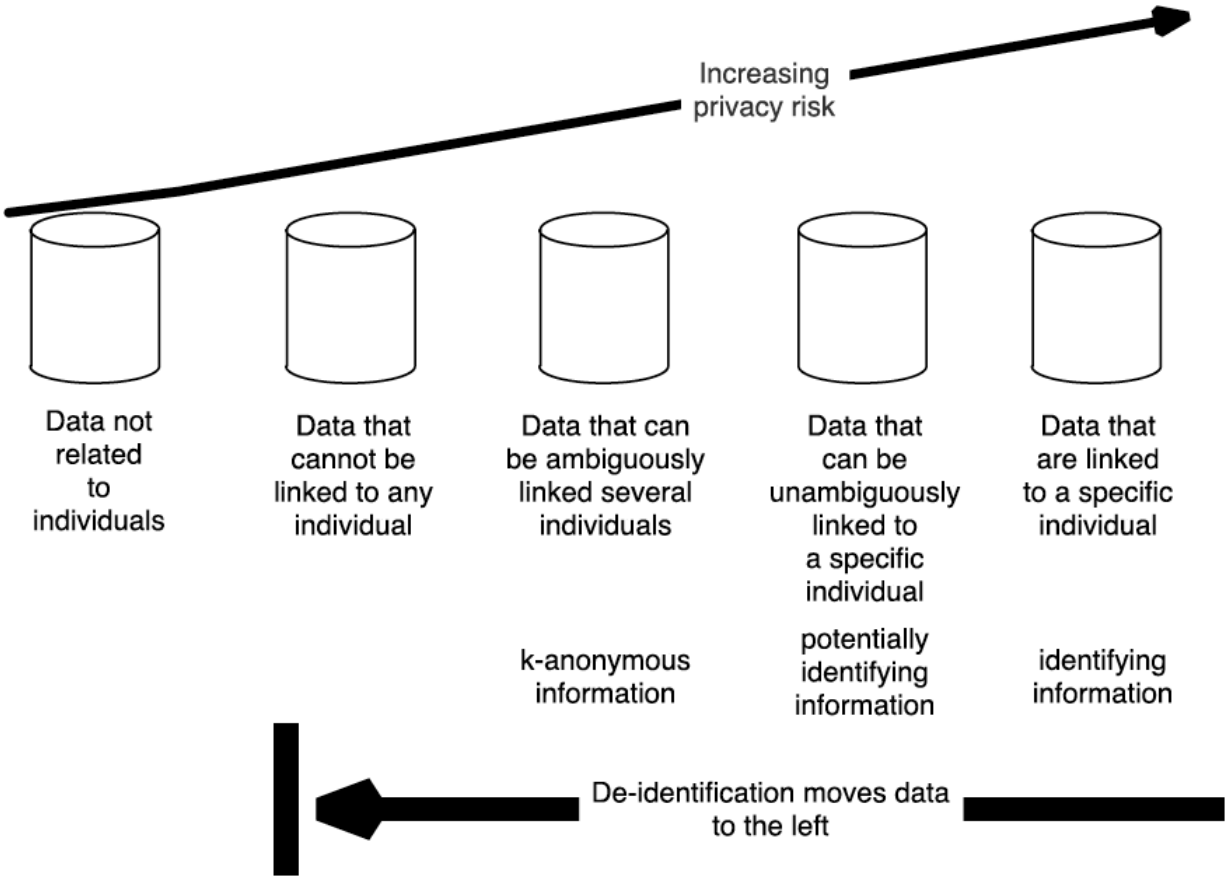


Figure 1 of the NIST De-Id Report

Anonos BigPrivacy separates sensitive or identifying data into segments and dereferences these segments using DDID pointers that obscure the identities of, and the relationships between and among, segmented data elements. Security and privacy of data is thereby improved by dynamically controlling levels of de-identification along a spectrum of de-identification as illustrated in Figure 1 of the NIST De-Id Report (copied above).

Anonosized data is decoded under controlled conditions to support certain uses within designated contexts as authorized by an individual data subject or by a Trusted Party. BigPrivacy retains the full capability to reproduce up to 100% of the original value and

utility of data, but it only authorizes that level of identifying information necessary to support each designated use, thereby enforcing appropriate levels of necessity and proportionality. BigPrivacy controls “identification” and “association” of data elements so data uses are isolated to those properly permissioned by means of JITI keys that provide context and meaning for DDIDs in accordance with data protection policies. If new authorized data uses arise, up to 100% of original data value and utility may be retained to support them.

Anonos BigPrivacy transforms data by leveraging:

1. **Segmentation:**
2. **Dereferencing; and**
3. **Dynamism**

- **<u>Segmentation</u> – sensitive or identifying data is separated into subsets.**

record_id	record_date	name	bpm	address	city	state	country	birth_date	death_date	home_tel
96325821	2008-11-14	e Freemont	55	4 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96502837	2011-01-23	Jane Freemont	59	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96627858	2011-11-12	Jane Freemont	62	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96520526	2011-03-08	Fred Smith	68	2091 Sycamore Pl	New Haven	CT	USA	1949-06-...	<null>	2030518892
96634010	2012-01-13	Fred Smith	72	2091 Sycamore Pl	New Haven	CT	USA	1949-06-...	2012-01-...	2030518892
96707533	2012-06-20	Stan Hart	53	33 E 8th St	Conroe	TX	USA	1960-01-...	<null>	7130864140
96682395	2012-03-27	Ginger Tu	71	12821 Acacia Ln	Sewickley	PA	USA	1958-09-...	<null>	4120438282
96840106	2013-02-26	Corrine Delacroix	<null>	491 Rue de Laxe	Paris	<null>	FRA	1959-03-...	<null>	3319058221...

In the above example, segmentation is accomplished by separating each data record into subsets comprised of each columnar data type (e.g., name, beats per minute (bpm), address) within each row.

- **<u>Dereferencing</u> – data values are replaced by DDID pointers, each of which provide access to the value to which the DDID points under controlled conditions.**

record_id	record_date	name	bpm	address	city	state	country	birth_date	death_date	home_tel
96325821	200[redacted]	Jane Freemont	55	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96502837	2011-01-23	Jane Freemont	59	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96627858	2011-11-12	Jane Freemont	62	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96520526	2011-03-08	Fred Smith	68	2091 Sycamore Pl	New Haven	CT	USA	1949-06-...	<null>	2030518892
96634010	2012-01-13	Fred Smith	72	2091 Sycamore Pl	New Haven	CT	USA	1949-06-...	2012-01-...	2030518892
96707533	2012-06-20	Stan Hart	53	33 E 8th St	Conroe	TX	USA	1960-01-...	<null>	7130864140
96682395	2012-03-27	Ginger Tu	71	12821 Acacia Ln	Sewickley	PA	USA	1958-09-...	<null>	4120438282
96840106	2013-02-26	Corrine Delacroix	<null>	491 Rue de Laxe	Paris	<null>	FRA	1959-03-...	<null>	3319058221...

record_id	record_date	name	bpm	address	city	state	country	birth_date	death_date	home_tel
96325821	200[redacted]	RD-b19fb7de	AD-4a7e8d33	RD-503a808c	RD-adcf63e7	RD-5f89d996	USA	RD-f2148b47	RD-0e9b97c3	RD-1539d067
96502837	2011-01-23	RD-9215622c	AD-4a7e8d33	RD-f2ddbf79	RD-2b607b8f	RD-46c3a855	USA	RD-4f0b03c0	RD-53ff450d	RD-a1ca9fb6
96627858	2011-11-12	RD-cdba5e16	AD-06e8eb04	RD-bf183476	RD-7be4fd4a	RD-f00d1700	USA	RD-f0b4b0d9	RD-9e4ade09	RD-5d429d7f
96520526	2011-03-08	RD-2e034a66	AD-06e8eb04	RD-d3b555c2	RD-5220435b	RD-917acc8d	USA	RD-40a668b0	RD-99fe3910	RD-6e9d43c5
96634010	2012-01-13	RD-528cc520	AD-2e431249	RD-3ceb28dc	RD-3fa2452a	RD-8e2983e3	USA	RD-2f628f8d	RD-f887a9b9	RD-a82d4f52
96707533	2012-06-20	RD-5e7e8faf	AD-4a7e8d33	RD-51a3e27e	RD-35172b7f	RD-7bad5db8	USA	RD-9ac379e5	RD-0ab48f45	RD-c9f57618
96682395	2012-03-27	RD-904e8477	AD-2e431249	RD-0dd51312	RD-8461b50c	RD-58cbabfb	USA	RD-529fb95c	RD-1d520617	RD-c734fb86
96840106	2013-02-26	RD-1105fdf9	AD-06cf9aeb	RD-b7d97f50	RD-422648ab	RD-25b032d9	FRA	RD-2cc750ff	RD-1e3bfe5f	RD-029f3473

Dereferencing in the above example is represented by replacing the first occurrence of "Jane Freemont" with the DDID pointer "RD-b19fb7de".

- **<u>Dynamism</u> – using different DDID pointers to replace different occurrences of the same data value.**

record_id	record_date	name	bpm	address	city	state	country	birth_date	death_date	home_tel
96325821	200[redacted]	Jane Freemont	55	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96502837	2011-01-23	Jane Freemont	[redacted]	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96627858	201[redacted]	Jane Freemont	62	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96520526	2011-03-08	Fred Smith	68	2091 Sycamore Pl	New Haven	CT	USA	1949-06-...	<null>	2030518892
96634010	2012-01-13	Fred Smith	72	2091 Sycamore Pl	New Haven	CT	USA	1949-06-...	2012-01-...	2030518892
96707533	2012-06-20	Stan Hart	53	33 E 8th St	Conroe	TX	USA	1960-01-...	<null>	7130864140
96682395	2012-03-27	Ginger Tu	71	12821 Acacia Ln	Sewickley	PA	USA	1958-09-...	<null>	4120438282
96840106	2013-02-26	Corrine Delacroix	<null>	491 Rue de Laxe	Paris	<null>	FRA	1959-03-...	<null>	3319058221...

record_id	record_date	name	bpm	address	city	state	country	birth_date	death_date	home_tel
96325821	200[redacted]	RD-b19fb7de	AD-4a7e8d33	RD-503a808c	RD-adcf63e7	RD-5f89d996	USA	RD-f2148b47	RD-0e9b97c3	RD-1539d067
96502837	2011-01-23	RD-9215622c	[redacted]e8d33	RD-f2ddbf79	RD-2b607b8f	RD-46c3a855	USA	RD-4f0b03c0	RD-53ff450d	RD-a1ca9fb6
96627858	201[redacted]	RD-cdba5e16	AD-06e8eb04	RD-bf183476	RD-7be4fd4a	RD-f00d1700	USA	RD-f0b4b0d9	RD-9e4ade09	RD-5d429d7f
96520526	2011-03-08	RD-2e034a66	AD-06e8eb04	RD-d3b555c2	RD-5220435b	RD-917acc8d	USA	RD-40a668b0	RD-99fe3910	RD-6e9d43c5
96634010	2012-01-13	RD-528cc520	AD-2e431249	RD-3ceb28dc	RD-3fa2452a	RD-8e2983e3	USA	RD-2f628f8d	RD-f887a9b9	RD-a82d4f52
96707533	2012-06-20	RD-5e7e8faf	AD-4a7e8d33	RD-51a3e27e	RD-35172b7f	RD-7bad5db8	USA	RD-9ac379e5	RD-0ab48f45	RD-c9f57618
96682395	2012-03-27	RD-904e8477	AD-2e431249	RD-0dd51312	RD-8461b50c	RD-58cbabfb	USA	RD-529fb95c	RD-1d520617	RD-c734fb86
96840106	2013-02-26	RD-1105fdf9	AD-06cf9aeb	RD-b7d97f50	RD-422648ab	RD-25b032d9	FRA	RD-2cc750ff	RD-1e3bfe5f	RD-029f3473

Dynamism in the above example is represented by replacing each occurrence of "Jane Freemont" with a different DDID pointer – i.e., "RD-b19fb7de", "RD-9215622c" and "RD-cdba5e16".

Research by Latanya Sweeney and Phillip Golle shows that knowledge of a birthdate, gender and zip code can be enough to identify as many as 62% - 87% of the people in the United States. However, in order to combine a birthdate, gender and zip code to achieve this 62% - 87% rate of re-identification, these three pieces of information must

be known *a priori* to relate to the same individual. By associating a different dynamically changing DDID pointer with each of birthdate, gender and zip code, it could not be known *a priori* if a birthdate, gender or zip code related to the same person or to some combination of different people. In this manner, ***Anonos uniquely helps to defeat the Mosaic Effect – i.e., without BigPrivacy, the more data sources that exist, the easier it becomes to unearth an individual data subject's identity.***

Each BigPrivacy pointer is comprised of a dynamically changing[24] 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 associated with data protection policies that provide context and meaning for each DDID pointer under controlled conditions.

The original value of each designated primary or indirect (quasi-) identifier is replaced by two types of DDIDs depending on the type of dereferencing desired:

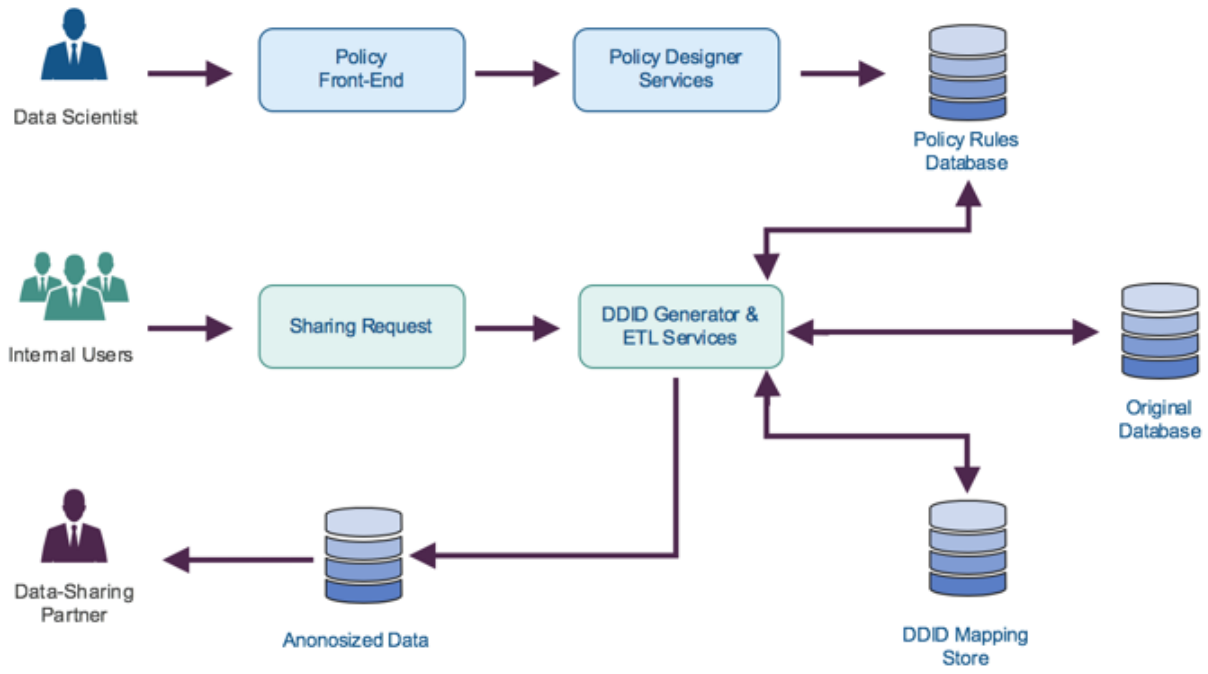
- **Identity Dereferencing** – where a DDID used to replace a data element is intended to point to the value of the replaced data element via a JITI key, the DDID is referred to as a **Replacement DDID** or **R-DDID**.

- **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**.

[24] 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.

R-DDIDs and A-DDIDs may dynamically change and may be temporally unique[25] when used for a different analyses or purpose.

BigPrivacy provides localized, technology-enforced policies for controlling sharing of Anonosized data in a dynamically de-identified/anonymous format. Access to perturbed or original versions of data is controlled via JITI keys in accordance with policies.



BigPrivacy Architecture Modeled After HIPAA Expert Determination Method

Development of Anonos BigPrivacy technology was modeled after de-identification requirements under HIPAA.[26] While BigPrivacy has capabilities that go beyond HIPAA requirements, it was essential to use HIPAA as a key model. This is because HIPAA embodies years of expertise dealing with removing and obscuring protected health

[25] The phrase "temporally unique" means that the time of initial assignment of a DDID to a data subject, the action, activity, process or trait is known, but the time period of assignment may be of any duration, limited or even perpetual.
[26] Under HIPAA, de-identified data is health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual. See http://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

information. A book published in March 2016 entitled *Using Health Information Technology to Improve Processes and Outcomes in Cancer*[27] notes:

> *The ability to convert data into information and then into knowledge is a distinct area of expertise and a scientific domain, called "data science." Fundamental shifts are occurring in our ability to organize, manage, annotate, and learn from data.* ***The most visible of these shifts is an increasing emphasis on big data and the concomitant transformation of each person in our society from being a data consumer to being a data provider.*** (emphasis added)

In the context of healthcare, BigPrivacy's principal goal is to technologically empower data scientists at the heart of de-identification efforts to move beyond the "Middle Ages" approach of repeated, bespoke manual assessments – analogous to monks' manually copying manuscripts – to better leverage and scale the data scientists' expertise. 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.[28] To support Information Age healthcare, Anonos believes that granularized data protection must be embedded into data at the data element level in order to reduce the risk of re-identification from the increasing volume, variety and velocity of data in today's data-driven society. ***BigPrivacy's patented systems and methods for granularized data protection overcome the inabilities of traditional nonintegrated security and privacy, thus defeating the Mosaic Effect and supporting the maximum extraction of value from data.***

The Big Data deluge transforming healthcare makes it difficult for data scientists to keep up with demand. BigPrivacy technology makes it much easier by leveraging scarce data

[27] http://www.amazon.com/Oncology-Informatics-Information-Technology-Processes/dp/0128021152
[28] See *supra* note 11.

scientist expertise through the creation of libraries of use case-specific policies that expressly delineate data fields comprising primary identifiers and indirect (quasi-) identifiers – along with programmatic instructions for transforming specified data fields to technologically enforce said policies. By employing generally accepted statistical and scientific principles and methods for rendering information not individually identifiable, policies can be established such that:

i. The risk is very small that information which is technologically enforced via each policy could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and

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

Technologically enforcing policies will help to reduce the likelihood of re-identification to acceptable (or better-than-acceptable) levels; libraries of policies will help support de-identification on a scalable basis.

- **Primary Identifiers**

BigPrivacy replaces primary identifiers with Replacement Dynamic De-Identifiers ("R-DDIDs") that are randomly assigned; correlations between R-DDIDs and the primary identifiers they replace are evident only with the use of JITI keys under the control of data subjects or Trusted Parties. The combination of these actions helps to defeat the Mosaic Effect while retaining access for up to 100% of data value and utility under controlled conditions.

record_id	record_date	name	bpm	address	city	state	country	birth_date	death_date	home_tel
96325821	200[redacted]	Jane Freemont	55	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96502837	2011-01-23	Jane Freemont	59	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96627858	2011-11-12	Jane Freemont	62	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96520526	2011-03-08	Fred Smith	68	2091 Sycamore Pl	New Haven	CT	USA	1949-06-...	<null>	2030518892
96634010	2012-01-13	Fred Smith	72	2091 Sycamore Pl	New Haven	CT	USA	1949-06-...	2012-01-...	2030518892
96707533	2012-06-20	Stan Hart	53	33 E 8th St	Conroe	TX	USA	1960-01-...	<null>	7130864140
96682395	2012-03-27	Ginger Tu	71	12821 Acacia Ln	Sewickley	PA	USA	1958-09-...	<null>	4120438282
96840106	2013-02-26	Corrine Delacroix	<null>	491 Rue de Laxe	Paris	<null>	FRA	1959-03-...	<null>	3319058221...

record_id	record_date	name	bpm	address	city	state	country	birth_date	death_date	home_tel
96325821	200[redacted]	RD-b19fb7de	AD-4a7e8d33	RD-503a808c	RD-adcf63e7	RD-5f89d996	USA	RD-f2148b47	RD-0e9b97c3	RD-1539d067
96502837	2011-01-23	RD-9215622c	AD-4a7e8d33	RD-f2ddbf79	RD-2b607b8f	RD-46c3a855	USA	RD-4f0b03c0	RD-53ff450d	RD-a1ca9fb6
96627858	2011-11-12	RD-cdba5e16	AD-06e8eb04	RD-bf183476	RD-7be4fd4a	RD-f00d1700	USA	RD-f0b4b0d9	RD-9e4ade09	RD-5d429d7f
96520526	2011-03-08	RD-2e034a66	AD-06e8eb04	RD-d3b555c2	RD-5220435b	RD-917acc8d	USA	RD-40a668b0	RD-99fe3910	RD-6e9d43c5
96634010	2012-01-13	RD-528cc520	AD-2e431249	RD-3ceb28dc	RD-3fa2452a	RD-8e2983e3	USA	RD-2f628f8d	RD-f887a9b9	RD-a82d4f52
96707533	2012-06-20	RD-5e7e8faf	AD-4a7e8d33	RD-51a3e27e	RD-35172b7f	RD-7bad5db8	USA	RD-9ac379e5	RD-0ab48f45	RD-c9f57618
96682395	2012-03-27	RD-904e8477	AD-2e431249	RD-0dd51312	RD-8461b50c	RD-58cbabfb	USA	RD-529fb95c	RD-1d520617	RD-c734fb86
96840106	2013-02-26	RD-1105fdf9	AD-06cf9aeb	RD-b7d97f50	RD-422648ab	RD-25b032d9	FRA	RD-2cc750ff	RD-1e3bfe5f	RD-029f3473

In the above example, the first occurrence of primary identifier "Jane Freemont" is replaced with the R-DDID "RD-b19fb7de". However, subsequent occurrences of "Jane Freemont" are replaced with different R-DDIDs to reduce the risk of unauthorized re-identification via the Mosaic Effect.

R-DDIDs, by way of encoding, may represent the de-identified version of any arbitrary data, including standard data like text and numeric values, as well as more exotic data – e.g., images, audio or video recordings. BigPrivacy can support optional R-DDID generation strategies, such as retrieving R-DDIDs from a remote API and creating branded R-DDIDs by incorporating static identifiers specific to a business name, research grant, etc.

- **Indirect (Quasi-) Identifiers**

In addition to replacing indirect (quasi-) identifiers with R-DDIDs, BigPrivacy may also insert Association Dynamic De-Identifiers ("A-DDIDs"), each of which represents the association of a replaced indirect identifier with a designated correlation schema or cohort. As more fully described in Section 5.5 Unstructured Data/Cohort-Based Research below, with regard to quasi-identifiers replaced with A-DDIDs, each A-DDID used to replace a quasi-identifier reflects a specific correlation schema or cohort.

record_id	record_date	name	bpm	address	city	state	country	birth_date	death_date	home_tel
96325821	2008-11-14	Jane Freemon...	55	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96502837	2011-01-23	Jane Freemont	59	[illegible] St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96627858	2011-11-12	Jane Freemon...	62	904 32nd St N	Milwaukee	WI	USA	1944-10-...	<null>	4140247282
96520526	2011-03-08	Fred Smith	68	2091 Sycamore Pl	New Haven	CT	USA	1949-06-...	<null>	2030518892
96634010	2012-01-13	Fred Smith	72	2091 Sycamore Pl	New Haven	CT	USA	1949-06-...	2012-01-...	2030518892
96707533	2012-06-20	Stan Hart	53	33 E 8th St	Conroe	TX	USA	1960-01-...	<null>	7130864140
96682395	2012-03-27	Ginger Tu	71	12821 Acacia Ln	Sewickley	PA	USA	1958-09-...	<null>	4120438282
96840106	2013-02-26	Corrine Delacroix	<null>	491 Rue de Laxe	Paris	<null>	FRA	1959-03-...	<null>	3319058221...

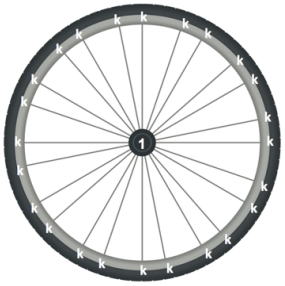
record_id	record_date	name	bpm	address	city	state	country	birth_date	death_date	home_tel
96325821	2008-11-14	RD-b1...	AD-4a7e8d33	RD-503a808c	RD-adcf63e7	RD-5f89d996	USA	RD-f2148b47	RD-0e9b97c3	RD-1539d067
96502837	2011-01-23	RD-9215622c	AD-4a7e8d33	[illegible]bf79	RD-2b607b8f	RD-46c3a855	USA	RD-4f0b03c0	RD-53ff450d	RD-a1ca9fb6
96627858	2011-11-12	RD-cd...	AD-06e8eb04	RD-bf183476	RD-7be4fd4a	RD-f00d1700	USA	RD-f0b4b0d9	RD-9e4ade09	RD-5d429d7f
96520526	2011-03-08	RD-2e034a66	AD-06e8eb04	RD-d3b555c2	RD-5220435b	RD-917acc8d	USA	RD-40a668b0	RD-99fe3910	RD-6e9d43c5
96634010	2012-01-13	RD-528cc520	AD-2e431249	RD-3ceb28dc	RD-3fa2452a	RD-8e2983e3	USA	RD-2f628f8d	RD-f887a9b9	RD-a82d4f52
96707533	2012-06-20	RD-5e7e8faf	AD-4a7e8d33	RD-51a3e27e	RD-35172b7f	RD-7bad5db8	USA	RD-9ac379e5	RD-0ab48f45	RD-c9f57618
96682395	2012-03-27	RD-904e8477	AD-2e431249	RD-0dd51312	RD-8461b50c	RD-58cbabfb	USA	RD-529fb95c	RD-1d520617	RD-c734fb86
96840106	2013-02-26	RD-1105fdf9	AD-06cf9aeb	RD-b7d97f50	RD-422648ab	RD-25b032d9	FRA	RD-2cc750ff	RD-1e3bfe5f	RD-029f3473

In the above example, the first heart rate (or bpm) of "55" and the second of "59" are replaced with the same A-DDID "AD-4a7e8d33", as they represent the correlation schema or cohort for heart rates between 51 and 60 beats per minute; the third heart rate of "62" is replaced with A-DDID "AD06e8eb04", as it represents the correlation schema or cohort for heart rates between 61 and 70 beats per minute.

BigPrivacy Enhances Data Accuracy and Use

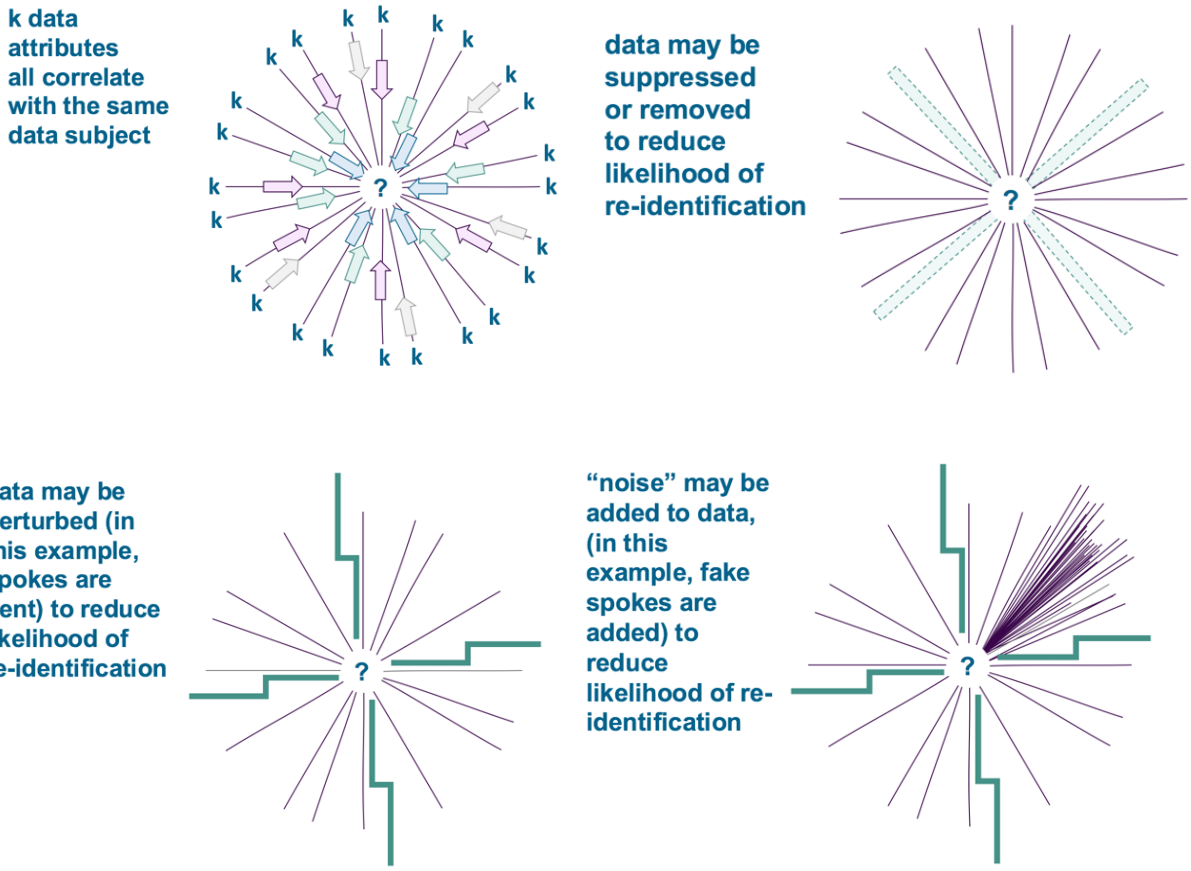
The following illustrations use a bicycle wheel to represent a data set unprotected by BigPrivacy, with the hub representing a Data Subject ("DS") and each spoke representing a quasi-identifier or Data Attribute ("DA") for the DS. To protect the identity of a DS, the data set may be encrypted. However, the data set will not be usable when encrypted; and when decrypted to enable use, the identity of the DS "hub" and all "spoke" DAs will be revealed and vulnerable. This is because traditional approaches to data security are premised on a binary "on/off" or "encrypt to protect/decrypt to use" model.

Source data =
1 data subject
k data attributes

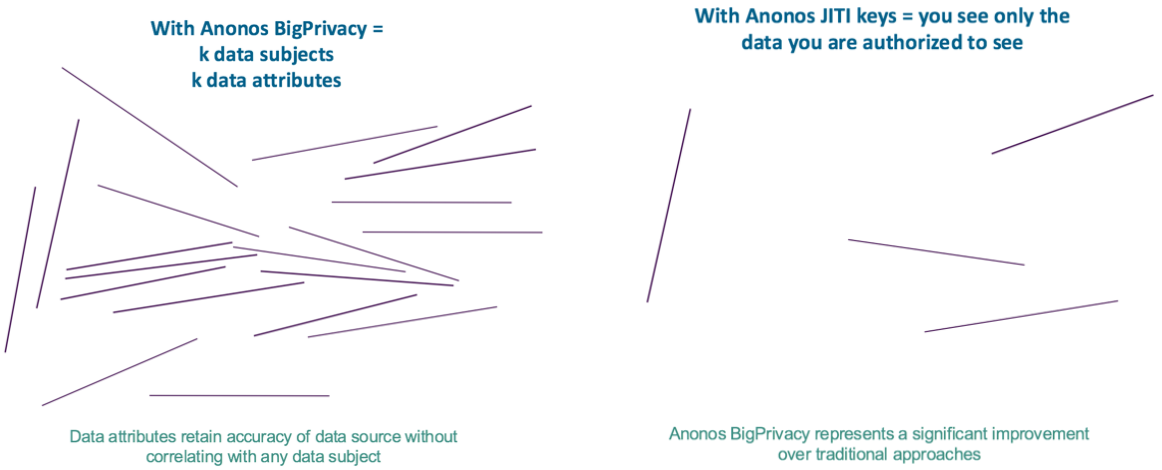


Encrypted data without key =
You see nothing

Even if the identity of the DS "hub" in the bicycle wheel is obscured, all "spoke" DAs still point to that hub, meaning there is a risk of re-identifying the DS. ***This explains why traditional static approaches to de-identification that use data suppression, perturbation, addition of noise, etc. to purposefully reduce the value and accuracy of "spoke" DAs are relying on a "1-to-k" model with "1" DS and "k" DAs, where <u>all DAs</u> are associated with the <u>same</u> DS.***



Conversely, Anonos supports a "k-to-k" model with "k" DSs associated with "k" DAs. This is analogous to removing spokes from a number of bicycle wheels and intermixing the spokes together. Secure reference tables provide information necessary to re-associate "spoke" DAs and appropriate "hub" DSs, but JITI keys (inaccessible to most and available only on a controlled need-to-know basis) are required to reveal this information. In this manner, associations between "spoke" DAs and "hub" DSs are not evident without access to JITI keys. Anonos BigPrivacy technology (which itself enhances security) therefore represents a significant improvement over using traditional encryption approaches alone – since, while they provide security benefits, they fail to address data privacy issues.



Anonos BigPrivacy breaks assumptions encoded in datasets, assumptions often used to attempt to achieve re-identification, all without reducing data value or utility. Anonosizing data introduces entropy at the most fundamental level – the DS level. What this means is that any given DS can map to any 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. Thus, 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 vitiate data value, but the "k-to-k" mapping we have described. This is entropy without the data value destruction that is very likely and, in many cases, certain to constrain scientific and research advances based on data relationship discovery.

Entropy injected into data sets by dereferencing identity (via replacing data identifiers with R-DDIDs) can help mitigate re-identification for data uses while enabling DAs to retain up to 100% accuracy, all without any visible associations with any DS. The ability to combine dereferencing identity with R-DDIDs and dereferencing associations with A-DDIDs to convey ranges or other associations/correlations/cohorts in a non-identifying manner decreases the need to distort, delete or otherwise vitiate the data in other data uses.

Certain countries (e.g., Denmark) are known for their medical research capabilities because the government collects data every time someone goes to the hospital; this data is then entered into a national administrative database. This kind of research is not currently possible in the U.S. because myriad siloed databases are used – there is no aggregated U.S. national database to query. Using Anonos BigPrivacy, multiple databases can be analyzed together where each is obscured using DDIDs and without having to access identifying versions of the underlying data. ***This enables analysis of combined data sets in a privacy-respectful manner, supporting faster, more in-depth, more data-driven research without "missing parts" that could resolve otherwise unanswerable questions about diseases, processes or cures.*** By enabling access to underlying granular data without revealing the identity of the data subject, privacy is protected while research advances, neither harming the other. Rather than providing a third party (often the researchers or data scientists) with access to personally identifying data, a dynamically obscured version of the data can be produced that provides individual data records which comply with a prearranged schema or cohorts so that they do not reveal personally identifying protected health information ("PHI").

A-DDIDs enable third parties to receive data that has been de-identified in accordance with HIPAA requirements. ***By agreeing in advance to common, temporally limited correlate schemata or cohorts that multiple Covered Entities will use at the same time for similar queries****,* third parties can legally access and use resulting

"Anonosized" data sets that contain information about the same people or about the same types of people; and complex, integrated relationships among data values can surface, enabling the discovery and/or refinement of medical procedures, genetic tests, pharmaceutical drug development, targeted cures such as monoclonal antibodies and more, substantially advancing the value and pace of medical science. Yet, none of this would be possible without access to the individual data sets unmasked but with the data subjects protected. BigPrivacy, by making this possible, ***thereby enables functional interoperability without requiring syntactic or semantic interoperability.***

5. Commercial Healthcare Applications

5.1 Health Information Exchanges

Growing patient-centric approaches to healthcare delivery, increasing Electronic Health Record ("EHR") adoption, federal incentives for "meaningful use" of certified EHR technology, and pressures to reduce healthcare costs are driving strong growth in a Health Information Exchange market projected to grow 50% within the next 4 years: from approximately $1.0 billion to $1.5 billion. However, data security and privacy concerns that require infrastructure investments to support health information exchange have restricted the growth of this market.[29]

As noted in a paper entitled *Protecting Patient Privacy: Strategies for Regulating Electronic Health Records Exchange* by the New York Civil Liberties Union[30]:

> *Easily shareable electronic records threaten patient privacy, and can lead to security breaches, misuse of information, and most importantly, loss of patient control over confidential and sensitive health information.* ***This threatens the confidential communication between doctors and***

[29] https://www.reportbuyer.com/product/3833624/
[30] http://www.nyclu.org/files/publications/nyclu_PatientPrivacy.pdf

***patients that has been a bedrock principle of modern medicine**.* *Confidentiality ensures that patients seek out care, and that they are open and honest with their providers. Fully informed by the totality of a patient's circumstances, providers can render the best care possible. Patients who fear a loss of control over their private medical information may lose faith in their doctor—and in the health care system. They may fail to share critical information with their treating providers or they may avoid treatment altogether.*

Guaranteeing confidentiality and patient control over sensitive health information is critical to the success of electronic health information exchange. Only with confidence that personal medical information will be shared in ways that benefit them and not cause them harm will patients fully engage in this promising technological advancement.

***National experts insist that the capacity to achieve granular segregation of patient health care information is a key goal—and a critical success factor—in the implementation of health information networks**.* (emphasis added)

Programmatic enforcement of data protection policies using BigPrivacy technology can enable granular access to patient health care data while supporting data security and privacy requirements necessary for successful Health Information Exchanges.[31]

5.2 Minimum Necessary Requirements

The HIPAA Minimum Necessary Rule[32] reads as follows:

[31] BigPrivacy can also assist in complying with HIPAA Security Rule technical standards for "access controls" requiring technical policies and procedures for electronic information systems that maintain electronic PHI, in order to allow access only to those persons or software authorized to have access rights.
[32] 45 CFR 164.502(b)(1).

When using or disclosing protected health information or when requesting protected health information from another covered entity or business associate, a covered entity or business associate must make reasonable efforts to limit protected health information to the minimum necessary to accomplish the intended purpose of the use, disclosure, or request.

In 2013, the HIPAA Omnibus Rule[33] expanded the Minimum Necessary Rule to apply to not only Covered Entities[34] but also to Business Associates.[35] Minimum Necessary Rule implementation guidelines specify obligations to comply with Covered Entities' Minimum Necessary policies so Business Associates must comply with their Covered Entities' Minimum Necessary policies.

Historically, an important safeguard for PHI has been role-based access to it. The HIPAA Privacy Rule includes among its Minimum Necessary Requirements that Covered Entities and Business Associates identify workforce members who need access to PHI, the PHI to which they need access, and any conditions on such access. The HIPAA Security Rule addresses role-based access to PHI in the form of safeguards that are both administrative (requiring policies and procedures for authorizing access to PHI) and technical (requiring technical policies and procedures for electronic information systems that maintain electronic PHI to allow access only to those persons or software authorized to have access rights). In order to help satisfy Minimum Necessary obligations under the HIPAA Privacy Rule and Security Rule, BigPrivacy JITI keys can technologically limit access to only those workforce members and software programs requiring access to PHI.

[33] Modifications to the HIPAA Privacy, Security, Enforcement, and Breach Notification Rules under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act; Other Modifications to the HIPAA Rules" ("Omnibus Rule"), 78 Fed. Reg. 5566 (Jan. 25, 2013).
[34] Covered Entities are defined in HIPAA as (1) health plans, (2) health care clearinghouses, and (3) health care providers who electronically transmit any health information in connection with transactions for which HHS has adopted standards.
[35] A Business Associate is defined in HIPAA as a person or entity that performs certain functions or activities that involve the use or disclosure of protected health information on behalf of, or provides services to, a Covered Entity.

5.3 Precision Medicine

Key to personalized Precision Medicine are tools that can support "prospective" medicine, health risk assessments, acquisition of a detailed family history, analysis of genomic information, and clinical decision support in an electronic record. Developing a clear value proposition for patients to share such sensitive data is necessary to realize a vision of broad data aggregation.[36]

In a 2012 Informatics workshop, Deven McGraw (then director of the Health Privacy Project at the Center for Democracy & Technology and currently Deputy Director for Health Information Privacy at HHS Office for Civil Rights) noted:

> *…the end goal of privacy is not privacy itself, but trust. The goal is to build a trusted, accountable ecosystem for using data in ways that help individuals, communities, and populations. Privacy rules are structured largely around tools such as patient consent and data minimization or de-identification. These tools are critically important, McGraw said, but they are not the end goal.* ***They are tools to be used to build trust, along with other tools.*** *It is also important to remember that* ***consumers and patients want their data to be protected, and they want medicine and health care to be advanced****. These competing interests need to be considered and balanced when developing privacy policies. McGraw also suggested that too much time is spent focusing just on the issue of consent in lieu of addressing other important privacy protections.* ***Consent is not the same as privacy. Consent ends up shifting the burden for protecting privacy to the patient.*** *That said, when surveyed, people often say that they want to be asked before their data are used for research purposes.* ***There are efforts now to obtain general consent for future research because it is not possible to define all of the potential research uses of the data being collected today, but this***

[36] See *supra* note 11.

does not lead to a meaningful and informed consent for the patient. (emphasis added).

In President Obama's 2015 State of the Union address, he announced a bold new initiative to revolutionize healthcare – the Precision Medicine Initiative ("PMI"). The goal of the Precision Medicine Initiative is to leverage advances in genomics and health information technology to accelerate biomedical discoveries. To be successful, the PMI will require new methods for managing, analyzing and ensuring the security and privacy of large data sets in order to develop a research cohort that will "engage a million or more Americans who volunteer to contribute their health data over many years to improve health outcomes, fuel the development of new treatments for disease, and catalyze a new era of data-based and more precise preventive care and medical treatment." The Precision Medicine Initiative Working Group has already issued recommendations for security and privacy of individual information, including establishing safeguards against unintended release of data.[37]

Anonos BigPrivacy can serve as part of the core of a trusted, accountable PMI ecosystem to support sustainable insights and discoveries that can help advance medicine and health. In the context of the Precision Medicine Initiative, BigPrivacy can:

- Maximize the authorized use of the research cohort comprised of Americans who volunteer their data;

- Facilitate compliance with and auditability against state-of-the-art technically enforced data security and privacy policies;

- Support granularized consent for future research by retaining health data in highly secure, privacy protected and technically obscured formats to support any later decisions by data subjects to exercise meaningful and informed consent by

[37] https://www.nih.gov/precision-medicine-initiative-cohort-program/precision-medicine-initiative-cohort-program-frequently-asked-questions

selectively un-obscuring specific elements of data in order to enable the data subjects to participate in desired projects; and

- Minimize data bias created when data subjects do not answer questions, or do not answer questions truthfully, because of discomfort with identifying uses of their personal data. Data subjects may be more willing to contribute accurate data to create representative data sets when they know data remains technologically obscured until they – and they alone (or HIPAA authorized legal designees) – decide to authorize selective disclosures of their data in the future.

The results of a recent study highlight the significant value of granularized consent enabled by BigPrivacy technology. A recent study of nearly 600,000 people found 13 surviving adults with genetic abnormalities from which most people die as children.[38] ***Each of these 13 individuals, therefore, represents 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 are unable to identify any of the 13 people.*** With Anonos BigPrivacy, their consent could have been dynamic and granular so the data subjects could have initially granted consent to use only certain data but later would have had the flexibility to authorize that additional data be revealed in the service of developing breakthrough treatments or cures.

5.4 Data Breaches

The HIPAA Breach Notification Rule[39] requires Covered Entities and Business Associates to provide notification following a breach of unsecured PHI. Covered Entities and Business Associates may also be subject to penalties for breach-related noncompliance based on their level of negligence, the potential for class action lawsuits by aggrieved patients and in certain circumstances, personal or even criminal liability. Covered Entities and Business Associates sometimes encrypt data to avoid data breach

[38] http://www.nature.com/nbt/journal/v34/n5/full/nbt.3514.html
[39] 45 CFR §§ 164.400-414.

notification obligations, however, encryption protects only *data at rest* or *data in transit*. Significantly, encryption does not and cannot *protect data in use*. But data is most vulnerable to misuse, abuse and attack precisely *when it is in use*. This is the major gap filled by BigPrivacy technology, because the primary purpose of storing or transmitting data is to be able to use it. BigPrivacy enables data to achieve its best and highest purpose: *to be used*.

Specifically, by enforcing Data Protection by Default principles[40] via granularized data protection, BigPrivacy technology renders data unusable, unreadable, or indecipherable in the hands of unauthorized parties *while the data is in use*. Anonos BigPrivacy provides an incentive to Covered Entities and Business Associates to offer greater protection for data not only while in transit or at rest but also *when data is in use and, as noted above, most vulnerable to misuse, abuse and attack*. In addition, BigPrivacy assists in complying with state[41] and international (e.g., GDPR) data breach requirements.

5.5 Unstructured Data/Cohort-Based Research

The term "unstructured data" refers to information that either does not have a pre-defined data model or is not organized in a pre-defined manner. Unstructured data is typically text-heavy, but may contain data such as dates, numbers, and facts as well. Unstructured data also includes multimedia data, such as pictures, audio, videos, and the like. IBM estimates that unstructured data accounts for 80% of all information in organizations.[42]

Electronic Medical Records (EMRs) contain not only structured data (e.g., patient name, red blood cell count, blood pressure, ICD-disease codes, etc.) but also "notes" fields,

[40] See *supra* note 3.
[41] Forty-seven states, the District of Columbia, Guam, Puerto Rico and the Virgin Islands have enacted legislation requiring private or governmental entities to notify individuals of security breaches of information involving personally identifiable information.
[42] https://www.ibm.com/blogs/business-analytics/data-is-everywhere/

composed of unstructured data. Anonosizing such notes fields de-identifies the fields into R-DDIDs to protect patient privacy. Such R-DDIDs by themselves do not reveal any information contained in the notes fields, but provide means of retrieving the notes fields under controlled authorized conditions.

Anonosizing data by also using A-DDIDs enables cohorts to be identified in connection with structured as well as unstructured data. While an A-DDID may be associated with a range in a structured EMR data field (e.g., systolic blood pressure > 140 and < 160), an A-DDID representing a cohort value may also be associated with a particular condition described in an EMR unstructured data notes field using heuristics and artificial intelligence. Beyond these applications, multimedia forms of unstructured data, such as the outputs of MRI, CT, Positron Emission Tomography, ultrasound scans, and other procedures will also benefit from using A-DDIDs to extract information into cohorts. A-DDIDs de-identify cohorts obtainable from extractable data to present information in a manner that is not re-identifiable back to data subjects because cohorts and the data values associated with the cohorts can be used independent from the identities of data subjects. This increases information available to researchers since A-DDIDs may be used to designate cohort values that maximize data value without jeopardizing patient privacy.

5.6 Blue Button Initiative and Patient Reported Outcomes

The Blue Button Initiative[43] aims to enable patients to view their own personal health records online and download them. Several Federal agencies, including the Departments of Defense, Health and Human Services, and of Veterans Affairs, implemented this capability for their beneficiaries. In addition, Blue Button has pledges of support from numerous health plans and some vendors of personal health record vendors across the United States. Data from Blue Button-enabled sites can be used to create portable medical histories that facilitate dialogue among health care providers, caregivers, and other trusted individuals or entities.

[43] https://www.healthit.gov/patients-families/your-health-data

Combining the data rich potential Patient Reported Outcomes[44] and the Blue Button Initiative can open up whole new avenues for research, medical breakthroughs, personalized precision medicine, etc. However, this is predicated on more effective data privacy and security for PHI. By limiting who can see data and what they can do with it, outdated data security and privacy techniques work against Big Data goals of increasing data sharing and value creation. Rather than having privacy policies and procedures stand alone, Anonos BigPrivacy implements them as technologically enforced, scalable privacy-by-design principles fused with the data. The result of fusing mathematical, technical and policy controls into data is that you can share more data and make more use of it over the full life cycle of data.

As soon as PHI is sent to a third party at a patient's request, it is no longer protected by HIPAA and there is no way to retrieve or protect it. Health data is a favorite target for misuse, abuse and attack since it sells for a premium on the black market due to the highly personalized and valuable nature of the information it contains. The combination of government incentives to accelerate adoption of EHRs and extended access to medical records by numerous types of organizations in support of integrated healthcare increases the vulnerability of PHI and potential for breaches.[45]

Government incentives to adopt EHRs and extended access to EHRs increase the vulnerability of data; and encryption only protects data when it is at rest or in transit. ***Conversely, as described above, Anonos BigPrivacy protects data while in use.*** Combining encryption to protect data at rest and in transit with Anonos BigPrivacy to protect data in use provides greater protection for data at all times as well as when it is most vulnerable to misuse, abuse and attack. Without enhanced safeguards, programs like Patient Reported Outcomes and the Blue Button Initiative present open invitations to "bad guys" to engage in even more fraud and abuse. Social engineering (e.g.,

[44] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3227331/
[45] http://www.brookings.edu/research/papers/2016/05/05-health-care-data-breaches-yaraghi

phishing) is already used today to get social security numbers and other personal information – and even greater harm could be done from misappropriated PHI.

5.7 Revenue Cycle Management

With cash flows declining, margins tightening, and bad debt increasing, it's more important than ever for Covered Entities to maximize cash flow and revenue management. Audits and denied claims resulting from submitted claims not meeting specific payer requirements are costly and disrupt cash flows. Revenue Cycle Management ("RCM") systems enable Covered Entities to maximize cash collection, improve payer performance, reduce collection times and minimize denial of submitted claims.

De-identified data created with BigPrivacy technology can enhance RCM systems, allowing them to create a more accurate and robust corpus of benchmarking reimbursement data. With the shift from volume-based to value-based billing and personalized precision medicine, it is increasingly more important to ensure that the most accurate representation of treatments and services is captured in International Classification of Diseases ("ICD") coding. Doing this will maximize value and increase the speed and efficiency of turning submitted claims into cash.

5.8 Fraud and Abuse Mitigation

Over the past five years, the Centers for Medicare & Medicaid Services ("CMS") has worked with predictive analytics experts, data scientists, and law enforcement to identify and take action on cases of fraud, waste, and abuse ("FWA") in the Medicare program. CMS has reported that they have generated $1.5 billion in savings due to Big Data initiatives started in June 2011.[46]

[46] https://blog.cms.gov/2016/05/27/medicares-big-data-tools-fight-prevent-fraud-to-yield-over-1-5-billion-in-savings/

De-identified data created with BigPrivacy technology can enhance FWA systems by enabling a more accurate and robust corpus of benchmarking data to provide better insight into evaluating submitted claims and determining the appropriateness of ICD coding to combat FWA.

5.9 Data Minimization

BigPrivacy technology can reconcile tensions surrounding data minimization – the irreversible, wholesale deletion of data from the ecosystem for all time.

A principal goal of Anonos BigPrivacy is to provide technologically enforced control mechanisms that permit only authorized parties to access and use data – and then only for purposes that verified credentials authorize. Figure 4 represents the concept of a BigPrivacy-enabled secure environment (i.e., an Anonos "Circle of Trust" or "CoT®").

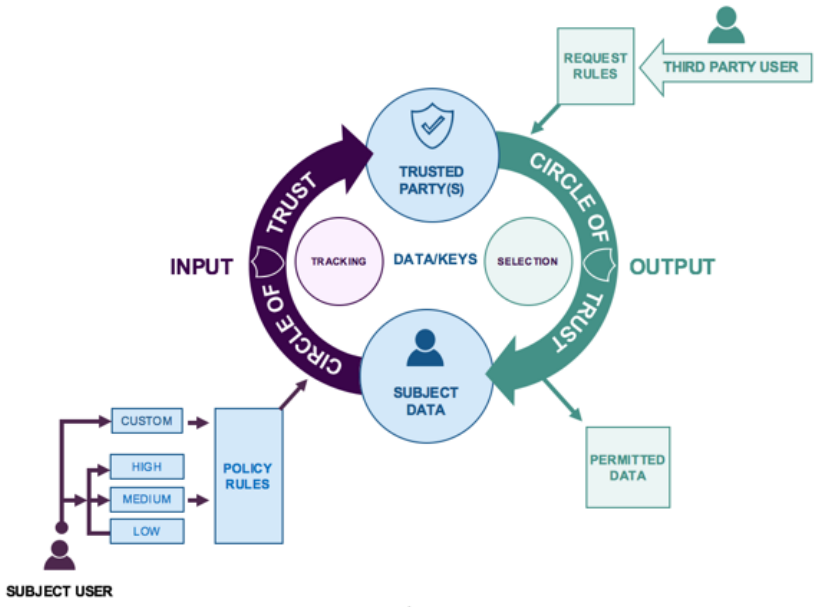


Figure 4

By supporting data <u>use</u> minimization without requiring <u>collection</u> minimization that otherwise removes data in perpetuity from the ecosystem for potential later authorized research, Anonos-enabled CoTs enable secondary uses of data without violating the privacy rights of data subjects.

The term “Tracking” in Figure 4 refers to tracking DDIDs used at different times for different purposes and the term “Selection” refers to selecting JITI keys necessary to correlate information represented by DDIDs with authorized purposes. Data may remain in original source databases in 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.

“Policy Rules” in Figure 4 relate to 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 “Anonosization” 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 other Privacy Enhancing Technologies (“PETs”) in connection with DDIDs, when to provide identifying information to facilitate transactions, etc. When data is input by someone other than the data subject to which data relates (a “Third Party User”), the Third Party User establishes “Request Rules” that enable data use/access in compliance with established corporate, legislative and/or regulatory data use/privacy requirements. “Permitted Data” represents data available for sharing with parties external to the CoT that satisfies Policy Rules established by the Subject User and/or Request Rules established by a Third Party User.

Note that there can be more than one Trusted Party authorized to work within a single Anonos-enabled CoT and that data subjects may participate in an unlimited number of CoTs. For increased security, CoTs can be implemented by means of centralized or federated models. Arrows in Figure 4 represent data movement; data inputs and outputs contain different information.

In approving a recent Federal Trade Commission (“FTC”) report, *Internet of Things – Privacy and Security in A Connected World,*[47] the majority of FTC Commissioners supported the need for data minimization, while a minority of FTC Commissioners noted

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

that such data minimization could negatively impact health research. BigPrivacy could help address the concerns of the majority of FTC Commissioners with regard to data use minimization without having to implement data collection minimization, while simultaneously addressing the concerns of the minority of FTC Commissioners that health research would be negatively impacted by data collection minimization. BigPrivacy technology can help minimize the risk of re-identification by ***protecting the privacy of individuals while enabling more complete data to be available for authorized use.***

In the EU, data privacy is a fundamental personal right. However, the absolute extreme of data privacy protection – e.g., data collection minimization – could usurp from individuals the right to make use of their own personal data. BigPrivacy technology can make available to them, at their election and control, the benefits of what their personal information can help to achieve – e.g., personalized precision medicine, advanced research, etc. Further, by minimizing the risk of identity disclosure while respecting and enforcing data protection for data subjects, BigPrivacy can enable them, again at their election and control, to avail themselves of potential benefits of authorized data use.

5.10 Cancer Moonshot/Genetic Research

As noted in the 2016 book *Oncology Informatics: Using Health Information Technology to Improve Processes and Outcomes in Cancer*[48]:

> *On January 12, 2016, the President of the United States upped the ante even further by announcing a "moon shot" for doubling the nation's progress against cancer over the next decade. As details of the Administration's efforts emerge,* ***it has become clear that a robust electronic infrastructure and improved policies for data sharing will be central to the moon shot efforts****.* (emphasis added)

[48] See *supra* note 11.

Anonos aims to help break down data silos to bring all the cancer fighters together – to work together, share information, and rapidly develop treatments for cancer.

- Anonos BigPrivacy is the only technology that can fundamentally remove significant technical barriers to the Cancer Moonshot with an innovative way for cancer centers to share cancer data without concerns that "it can't be done."

- BigPrivacy enables sharing of sensitive data without:
    - Requiring agreement on standards for collecting, labeling or processing; or
    - Compromising patient privacy.

- BigPrivacy does not require changes to existing systems.

- What is required is:

    - The establishment of a data harmonized format into which data will be transformed for analysis – all cancer centers can take data they are willing to share from existing systems in existing formats and transform the data into a harmonized format for purposes of analysis only.
    - BigPrivacy converts PHI into de-Identified data that is no longer subject to HIPAA – without compromising patient privacy. As data is de-identified, BigPrivacy simultaneously converts it into a harmonized format to enable analysis.

Anonos BigPrivacy technology overcomes concerns about patient privacy, improves cross-sectional and longitudinal analyses and defeats lack of interoperability among systems – all of which are necessary for a successful “Cancer Moonshot” initiative.[49]

[49] http://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative

An article from the January 2016 *Journal of Clinical Oncology* summarizes the challenges underlying the Cancer Moonshot:

> *Patients with cancer often receive care from a multitude of disparate sources, which include subspecialists, primary care physicians, ambulatory care offices, hospitals, laboratories, imaging facilities, and other health care facilities and organizations. Furthermore, most of these providers rely on EHR systems that lack interoperability and are cited for having poor usability and incomplete implementation of data standards, which can result in illogical and potentially dangerous alterations of physician workflow as well as high-cost/low-value outcomes.* ***In such a climate, patients with cancer inevitably will experience firsthand the fragmentation that characterizes the U.S. health care system and its current electronic infrastructure.***
>
> *This fragmented system of cancer care often hinders providers from gaining clinically meaningful, longitudinal insights into patient care and outcomes because the knowledge gained from individual patient encounters is rarely incorporated into larger data collection by multiple providers or health care systems. In addition, the limitations of the current clinical trial system restrict the ability to learn from the experiences of patients with cancer. Currently, only a tiny minority (reportedly <5% of adult patients with cancer in the United States) participate in clinical research studies during any part of their cancer care and are typically not representative of the population of patients with cancer. This suggests that an overwhelming majority of experiences of patients with cancer (>95%) do not contribute to the general oncology knowledge base.* ***As a consequence, few if any clinical insights are generated from these experiences to improve diagnosis and the treatment of future patients****.*[50] (emphasis added)

[50] http://jco.ascopubs.org/content/early/2016/01/07/JCO.2015.65.0598.full

CancerLinQ, an initiative of the American Society of Clinical Oncology ("ASCO"),[51] is a pioneering quality measurement and reporting system aimed at enabling oncologists to harness data for assessing, monitoring, and improving cancer care. To achieve its mission, CancerLinQ collects, aggregates, and analyzes data for cancer patients that originates in EHRs, practice management systems, or other data sources – which may exist in structured and/or unstructured format. CancerLinQ collects and uses PHI.[52]

While CancerLinQ has been successful in aggregating cancer-related data, its requirements underscore the challenges faced by the Cancer Moonshot. First, CancerLinQ requires "syntactic interoperability" – i.e., the ability of different health information systems that store the same type of information in different formats (i.e., heterogeneous formats) to exchange data. Second, it requires "semantic interoperability" – i.e., the ability of systems to understand the meaning of data exchanged; this requires shared data models, standard vocabularies and common data elements (for this purpose, CancerLinQ uses the NCI Metathesaurus as its unified medical language system). Third, because CancerLinQ handles PHI, it must have HIPAA compliant Business Associate arrangements in place with each Covered Entity with which it interacts. While CancerLinQ has had considerable success in its own right, ***it is unlikely that, in the near-term, all three of these requirements will be satisfied on a national scale.***

Anonos BigPrivacy can provide ***functional interoperability***, a way of enabling willing parties to participate in the Cancer Moonshot, all without requiring syntactic interoperability, semantic interoperability and/or Business Associate agreements between and among parties. Note that CancerLinQ still has to de-identify PHI before providing it to third parties outside of the CancerLinQ "private data enclave"; therefore, the advantages of using BigPrivacy DDIDs and JITI keys would still redound to CancerLinQ.

[51] Anonos has no association with ASCO or CancerLinQ.
[52] See *supra* note 47.

While CancerLinQ is a "private data enclave," the Cancer Genome Atlas ("TCGA"), a collaboration between the National Cancer Institute ("NCI") and National Human Genome Research Institute ("NHGRI"),[53] is an example of a public cancer data store with two data access tiers:

- **The Open Access Data Tier**, which comprises public data not unique to an individual and does not require user certification; and

- **The Controlled Access Data Tier**, which contains data that may be unique to an individual. All data types are stripped of direct identifiers. The Controlled Access data tier requires user certification.[54]

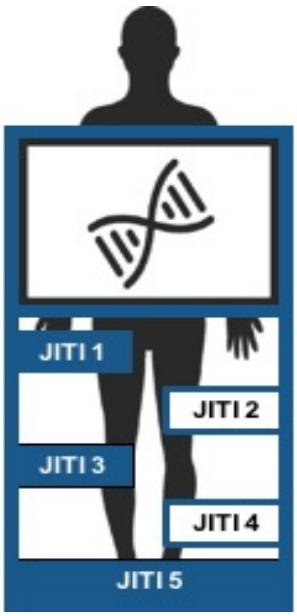
While two-tiered (i.e., open and controlled) data stores like the TCGA provide invaluable information (e.g., TCGA data has contributed to more than 1,000 studies of cancer by independent researchers), they present challenges when researchers desire access to covariate information from the controlled tier for the purpose of augmenting data from the open tier. While gaining access to the controlled tier requires user certification, once researchers have been certified, ***they gain access to the entire data set – significantly more data than required for the analysis at hand.*** This is because, as noted previously, traditional approaches to data security and privacy present binary options – yes or no, all or nothing – but they lack the means to support selective access to data. This often makes it necessary for an Institutional Review Board ("IRB") – a committee established to review and approve research involving human subjects to ensure research is conducted in accordance with federal, institutional, and ethical guidelines – to review a researcher's work to verify that the results of their analysis do not reveal (intentionally or unintentionally) any identifying element from the controlled data set provided to them.

[53] http://cancergenome.nih.gov/abouttcga/overview
[54] https://tcga-data.nci.nih.gov/tcga/tcgaAccessTiers.jsp

Conversely, BigPrivacy supports granularized access to identifiable data so that, if a researcher requests information from a controlled tier, that information is provided only to the extent of the minimum data necessary for their specifically requested analysis. In this manner, BigPrivacy can help enforce data access standards *analogous to those* required under the HIPAA Minimum Necessary Rule[55]:

> *"…[w]hen using or disclosing protected health information … reasonable efforts [should be taken] to* ***limit protected health information to the minimum necessary to accomplish the intended purpose of the use, disclosure, or request****."* (emphasis added)

Anonos BigPrivacy enables information to be revealed in successive gradations of precision via a technology enforced risk-based framework. In genomic research, this enables the relationship between a subject's phenotype (e.g. disease state) and genotype (their DNA) to be revealed in successive gradations of precision by revealing just the level of identifying information necessary for each level of authorized use.



- **JITI 1: Pathways bearing mutations and subjects in binary cohort groups**
- **JITI 2: JITI1 + Genes bearing mutations and detailed disease classification**
- **JITI 3: JITI2 + Specific gene variants and disease class scores**
- **JITI 4: JITI3 + Hapmap haplotype results and full disease history**
- **JITI 5: JITI 4 + Full SNP data and full patient record**

[55] See *supra* note 30.

By de-risking data via a risk-based framework that technologically enforces data protection policies in a contextually flexible, selective manner all the way down to lower data element levels, BigPrivacy technology can facilitate research by:

- Supporting a “minimum necessary” approach to data access, thereby enhancing the security and privacy of identifying human subject data;

- Reducing the work of IRBs in analyzing whether access to controlled tier data remains problematic since a much smaller subset of data will be revealed to researchers; and

- Saving time for researchers since controlled tier data will already be filtered down to the specific data necessary for their desired analysis.

6. BigPrivacy Glossary

A-DDID – the type of DDID used for Association Dereferencing.

Association Dereferencing – when a DDID is used to replace a data element 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.

Data Privacy – controlling which uses of data an authorized party is allowed to make.

Data Protection by Default – the most recent implementation of Privacy by Design (PbD) required under Recital 78 and Article 25 the EU General Data Protection Regulation (“GDPR”) that requires PbD techniques to 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 a specific product or service authorized by an individual data subject.

Data Security – preventing access to data by an unauthorized party.

DDIDs – Dynamic de-identifiers.

De-identification – the process of removing the association between a set of identifying data and the data subject.

De-risking Data - to severely minimize the likelihood of re-identification, without vitiating the research or other value of the underlying data.

FIPPS - Fair Information Practice Principles.

Identity Dereferencing – when a DDID is used to replace a data element to point to the value of the replaced data element via a JITI key.

JITI key – Just-In-Time-Identity key/schema necessary for transforming DDIDs into intelligible forms.

Mosaic Effect – a mathematical and statistical phenomenon that the more data sources that exist, the easier it becomes to unearth an individual data subject's identity.

Privacy by Design ("PbD") – the data privacy approach developed by Ann Cavoukian, Ph.D., former IPC Commissioner for embedding privacy into the system design process.

R-DDID – the type of DDID used for Identity Dereferencing.

Big Data in Healthcare and Life Sciences

Anonos® BigPrivacy® Technology Briefing

Jonas Almeida, Ph.D., is Professor and Chief Technology Officer at the Biomedical Informatics Department of Stony Brook University (State University of NY). His research is focused on the bioinformatics challenges to deploying personalized medicine solutions at the interface between cloud computing and machine learning. For a list of publications see pub.jonasalmeida.info.
https://www.linkedin.com/in/jonasalmeida

Sean Clouston, Ph.D., is Assistant Professor in the Department of Family, Population, and Preventive Medicine and Core Faculty in the Program in Public Health at Stony Brook University (State University of NY). His research focuses on causal analysis of longitudinal aging studies. For a list of publications,
see: ncbi.nlm.nih.gov/myncbi/browse/collection/4124.
https://www.linkedin.com/in/sean-clouston-a1850066

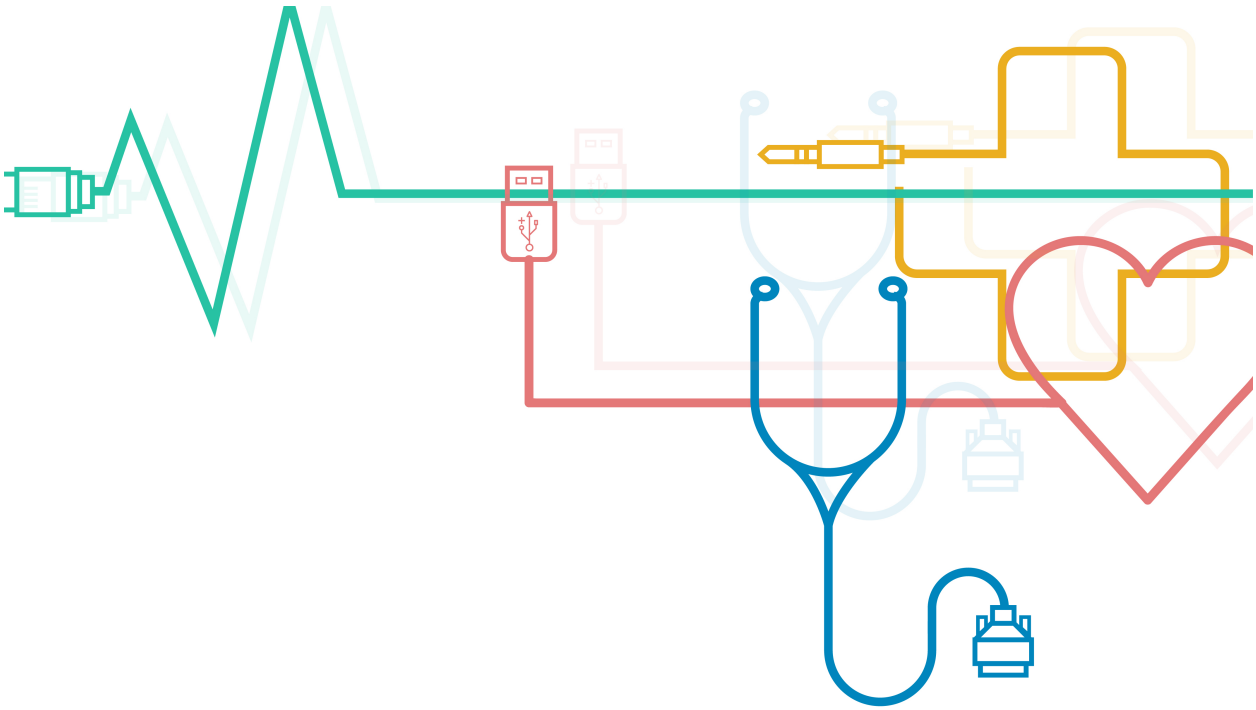
Gary LaFever is Co-Founder and Chief Executive Officer of Anonos Inc., developer of patented BigPrivacy technology that integrates security and privacy to maximize the value of restricted data by enabling sharing and repurposing of that data in a privacy respectful manner.
https://www.linkedin.com/in/garylafever

Ted Myerson is Co-Founder of Anonos Inc., developer of patented BigPrivacy technology that integrates security and privacy to maximize the value of restricted data by enabling sharing and repurposing of that data in a privacy respectful manner.
https://www.linkedin.com/in/tedmyerson

Sandeep Pulim, MD, is an advisor at HealthXL, and Chief Medical Information Officer at @Point of Care, a mobile clinician/patient collaboration platform, which combines clinician knowledge with patient data to achieve improved health outcomes. You will find Sandeep's thoughts on science, technology, healthcare, and medicine on Twitter at @spulim.
https://www.linkedin.com/in/spulim

Anonos® BigPrivacy Technology

Unlocks the Value of Health Big Data



If you have questions or comments, we would like to hear from you.

BigPrivacy@anonos.com