anonos®

Submitted to U.S. Federal Trade Commission at https://ftcpublic.commentworks.com/ftc/bigdataworkshop/ and to Republic of Mauritius Data Protection Office at pmo-dpo@mail.gov.mu

October 31, 2014

The Honorable Edith Ramirez, Chairwoman
Federal Trade Commission
600 Pennsylvania Avenue, N.W.
Washington, D.C. 20580

Mrs. Drudeisha Madhub, Commissioner
Data Protection Office
Emmanuel Anquetil Building
Port Louis, Republic of Mauritius

Re: (1) FTC Workshop - Big Data: A Tool for Inclusion or Exclusion *and*
(2) International Conference of Data Protection and Privacy Commissioners

Dear Chairwoman Ramirez and Commissioner Madhub:

We are writing this letter to:

- Express our appreciation to each of you for ongoing privacy protection efforts, specifically including the FTC Big Data Workshop, Project No. P145406: *A Tool for Inclusion or Exclusion* held in Washington, D.C. on September 15, 2014 (the "FTC Workshop"), which we participated at in person, and the *36th International Conference of Data Protection and Privacy Commissioners* held in Balaclava, Mauritius from October 13 to 16, 2014 (the "Privacy Conference"), which we viewed via live webcast; and

- Respectfully make observations and suggestions responsive to the purpose, intent, goals and objectives of the FTC Workshop and the Privacy Conference.

1. Language Can Sometimes Get in the Way

It was apparent at the Privacy Conference that real-time interpretation was necessary (and provided) to render oral communications from one language into another so that all participants, speaking a variety of languages, could benefit from the conference. Since English was the first language of the majority of participants at the FTC Workshop, it was less apparent that even common language sometimes gets in the way of effective communications when definitions differ. One exception was when a panelist at the FTC Workshop stated that "IP addresses are not personally identifying information - perhaps they should be, but they are not." The FTC Workshop panelist was referring to differing interpretations of what constitutes "personally identifiable information," "personal

information," or "PII". Commentators have noted that until interpretations of what constitutes PII are reconciled, "…*we will continue to lack a coherent approach to defining the proper boundaries of privacy regulation. Privacy law thus depends upon addressing the PII problem - it can no longer remain the unacknowledged elephant in the room*."[1] What constitutes PII is central to privacy regulation, however, despite the critical importance of this term there are varying interpretations and perspectives on the appropriate definition.

In order to avoid letting language get in the way of evaluating Dynamic Data Obscurity[2] as an approach to reconciling data privacy and value, we ask readers to "temporarily suspend" their opinions with respect to interpretations / meanings of terms used in this letter. Instead, we ask that readers accept definitions as provided / referenced in this letter on their face value for the sole purpose of assessing the promise of Dynamic Data Obscurity. For example, in this letter we use the term "Personal Data" or "PD" to refer to a data element that alone, or in combination with other data elements, can be linked or is linkable to an individual, which individual we refer to as a "Data Subject."

2. Anonymity is Not Dead – It Needs to Be Re-Imagined

In 2009, Professor Paul Ohm published a critical law review article entitled *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization* in which he argued that the concept of anonymity should be abandoned since so much information can be re-identified.[3] In 2010, the Federal Trade Commission (FTC) recognized problems with anonymity in the context of the ultimately ill-fated "Do Not Track" initiative.[4] Before the Committee on Energy and Commerce, Subcommittee on Commerce, Trade, and Consumer Protection of the US House of Representatives, the FTC acknowledged "…the blurring of the distinction between personally identifiable information and supposedly anonymous or de-identified information."[5]

---

[1] Schwartz, Paul M. and Solove, Daniel J., *The PII Problem: Privacy and a New Concept of Personally Identifiable Information* (December 5, 2011). New York University Law Review, Vol. 86, p. 1814. Available at http://ssrn.com/abstract=1909366

[2] See International Association of Privacy Professionals (IAPP) article on Dynamic Data Obscurity entitled *What Anonymization and the TSA Have in Common* at https://privacyassociation.org/news/a/what-anonymization-and-the-tsa-have-in-common/). See also Information Accountability Foundation blog entitled *Taking Accountability Controls to the Next Level: Dynamic Data Obscurity* at http://informationaccountability.org/taking-accountability-controls-to-the-next-level-dynamic-data-obscurity/.

[3] Ohm, Paul, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization* (August 13, 2009). UCLA Law Review, Vol. 57, p. 1701. Available at http://ssrn.com/abstract=1450006.

[4] Douglas Crawford notes in a 2014 article entitled *The Failure of 'Do Not Track,'* that Do not Track (DNT) "was a fantastic but ultimately idealistic idea that was always doomed to failure….DNT was proposed in response to US and EU legislators….demanding in 2007 that the internet advertising industry provide an agreed-upon standard which would allow consumers to opt-out of tracking by web advertisers….Unfortunately, the standard relies entirely on the cooperation of websites, advertisers, and analytics companies, who profit almost entirely from invading web users' privacy and tracking their actions across the web in order to deliver ever more individually targeted advertising." See https://www.bestvpn.com/blog/10854/the-failure-of-do-not-track/

[5] See http://www.ftc.gov/sites/default/files/documents/public_statements/prepared-statement-federal-trade-commission-do-not-track/101202donottrack.pdf

More recently, the 2014 President's Council of Advisors on Science and Technology report entitled *Big Data and Privacy: A Technological Perspective* (the "PCAST Report")[6] stated that "anonymization, while valuable in the past, [is] seen as having only limited future potential." And, the Article 29 Data Protection Working Party *Opinion 05/2014 on Anonymisation Techniques*[7] criticized anonymization techniques based on three criteria: (i) is it still possible to single out an individual, (ii) is it still possible to link records relating to an individual, and (iii) can information be inferred concerning an individual?

We contend that the above criticisms of anonymity were directed at static approaches to anonymity. The concept of Dynamic Data Obscurity generally, and the specific Anonos Dynamic Anonymity / Circle of Trust (CoT) approach to Dynamic Data Obscurity as described in this letter, provide a means to dynamically (versus statically) anonymize data thereby overcoming limitations of static anonymity and opening up the door to disprove the axiom that "You can have privacy or you can have value – but you cannot have both."

Woodrow Hartzog and Frederic Stutzman, in their law review article *The Case for Online Obscurity*,[8] touched upon the power of obscuring data when they asserted that:

> "On the Internet, obscure information has a minimal risk of being discovered or understood by unintended recipients. Empirical research demonstrates that Internet users rely on obscurity perhaps more than anything else to protect their privacy. Yet, online obscurity has been largely ignored by courts and lawmakers. In this Article, we argue that obscurity is a critical component of online privacy, but it has not been embraced by courts and lawmakers because it has never been adequately defined or conceptualized."

The illustrations below graphically illustrate, and build upon, a construct introduced by Jeffrey M. Skopek, in his law review article *Anonymity, the Production of Goods, and Institutional Design*[9] that anonymity (which he defines to encompass pseudonymity) provides an effective approach to balancing individual benefits from privacy and societal benefits from anonymity. *Since Dynamic Data Obscurity overcomes limitations of static anonymity, this construct provides an effective approach to evaluate the capabilities of Dynamic Data Obscurity generally, and the Anonos Dynamic Anonymity / Circle of Trust (CoT) approach to Dynamic Data Obscurity, to maximize both data privacy and data value.*

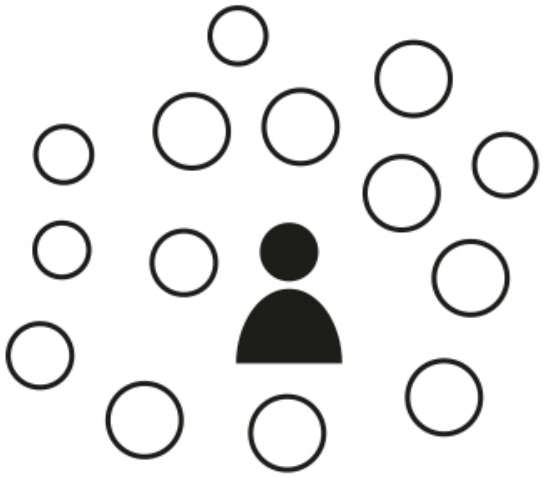
[6] See http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf
[7] See http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
[8] See https://www.futureofprivacy.org/wp-content/uploads/2011/07/The%20Case%20for%20Online%20Obscurity.pdf
[9] See http://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=4960&context=flr

Pure Condition of Privacy

In the pure condition of ***Privacy***, we have knowledge of a person's identity, but not of associated personal data.

Derived from Skopek, J. M. 2014. Anonymity, The Production Of Goods, And Institutional Design. Fordham Law Review 82(4): 1751-1809; http://ir.lawnet.fordham.edu/flr/vol82/iss4/4/.

Privacy benefits individual Data Subjects



Known Data Subject

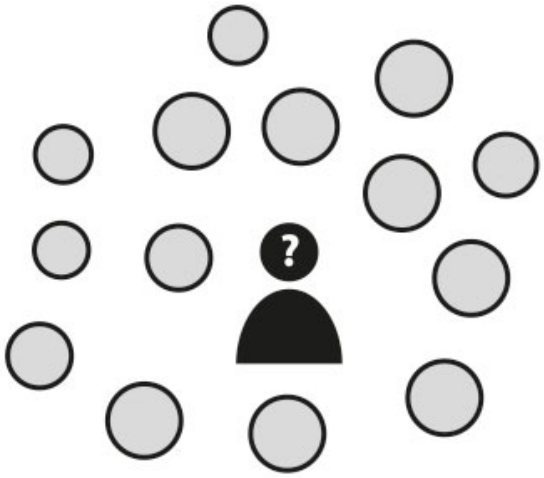
Unknown Personal Data

Pure Condition of Anonymity

In the pure condition of ***Anonymity***, we have knowledge of personal data, but not of the associated person's identity.

Derived from Skopek, J. M. 2014. Anonymity, The Production Of Goods, And Institutional Design. Fordham Law Review 82(4): 1751-1809; http://ir.lawnet.fordham.edu/flr/vol82/iss4/4/.

Anonymity benefits Society



Unknown Data Subject

Known Personal Data

Privacy vs. Anonymity

Privacy and anonymity are flip sides of each other – they function in opposite ways:

- ***Privacy*** can hide facts about someone whose identity is known by removing information and other goods associated with the person from public circulation.

- ***Anonymity*** can hide the identity of someone about whom facts are known for the purpose of putting such facts and related goods into public circulation.

Derived from Skopek, J. M. 2014. Anonymity, The Production Of Goods, And Institutional Design. Fordham Law Review 82(4): 1751-1809; http://ir.lawnet.fordham.edu/flr/vol82/iss4/4/.

But by dynamically obscuring relationships between data elements and Data Subjects, Anonos Dynamic Anonymity helps accomplish both:

- ***Enhanced privacy for data subjects with untrusted parties; and***

- ***Maximized value of data subject information for trusted parties.***

3. Unlocking the Potential of the Interconnected World.

When it comes to data, the idea that "You can have privacy or you can have value – but you cannot have both" is considered axiomatic. In fact, it is dangerously flawed. Zero privacy has three distinct disadvantages to data analytics. First, no privacy actually reduces the value of data because it does not filter out anyone or anything from the dataset, leaving too many choices and an excess of "noise." Second, no privacy subjects the identified or identifiable Data Subject associated with the data to potential discrimination and harm. Third, it imposes potential liability on data processors under regulations mandating information security and data protection requirements.

Providing privacy, on the other hand, has traditionally meant deleting data, usually through so-called anonymization practices. The distinct disadvantage of protecting privacy has been a reduction in available useful and relevant data with protection from re-identification. Attempts to protect privacy by using "static anonymity" (in which a single, unchanging identifier is used in an attempt to hide connections between data and Data Subject) also fail. Since static de-identification schemes can be easily broken, the expectation of privacy is never met.

As outlined in this letter, Anonos solves these difficult problems by using "Dynamic Anonymity" which creates dynamic de-identifiers (DDIDs). DDIDs can be further leveraged by dynamic data privacy controls within Anonos-enabled "Circles of Trust" (CoT) to maximize both data privacy and value.

4. A Promised Land?

We live at an unprecedented time in history, akin to standing on a mountaintop looking down at a "promised land" in the valley below, where ongoing improvements in technology make new societal benefits possible. In this promised land are a myriad of new products, offerings and services – including potential medical advances and personalized medicine derived from genetic research and personalized experiences made possible by the interconnected Internet of Things (IoT) / Internet of Everything (IoE). This promised land, however, remains fraught with risk.

Joel Kupersmith highlights some cautions relevant to genomic research in particular:[10]

> "The benefits of amassing genomic data in sufficient case numbers for validity and making this knowledge available to an appropriately wide body of expert investigators are extensive. Research derived from genomic databases offers potentially large health payoffs. Genomics can help scientists predict who will develop a disease (e.g., Huntington's disease) and tailor treatments. It also holds the potential to bring about a paradigm shift in how we think about and classify disease; i.e., allowing us to move from the pathology-based approach begun in the late 19th century – which focuses on the progression of disease in a specific organ – to a biochemical-and genomics-based approach. This new approach is already being applied to a number of diseases, including certain cancers….
>
> Yet the damage caused by the misuse of genomic data can be irreparable. Disclosure of genomic information can have stigmatizing consequences for both the individual and family, particularly first-degree relatives. These consequences include employment discrimination, denial of life insurance, and inappropriate marketing. And that is the conundrum: Realizing the promise of genomics, while safeguarding genomic data and protecting individual privacy….
>
> Until now, anonymization of data, database security, and these protective laws have been the primary safeguards against security lapses in genomic data made partially or largely available to the public. *However, various factors, including formation of very large databases, and data sharing and access by large numbers of individuals have put new strains on genomic security."* (Emphasis added)

[10] Kupersmith, Joel, *The Privacy Conundrum And Genomic Research: Re-Identification And Other Concerns*, The Health Affairs Blog at http://healthaffairs.org/blog/2013/09/11/the-privacy-conundrum-and-genomic-research-re-identification-and-other-concerns/

Some experts, like Plamen Nedeltchev, Ph.D., Distinguished IT Engineer at Cisco, believe the Internet of Things (IoT) / Internet of Everything (IoE) (where virtually every product, locale, personal item, and mobile device and entity has a unique IP address, making it remotely / wirelessly accessible) "is potentially the biggest business opportunity in the history of mankind" and that it "will change the world with extraordinary and wide-ranging implications, affecting everyone on the planet."[11] However, the IoT / IoE presents its own issues as noted below in an article entitled *The Creepy Factor of the Internet of Things*:[12]

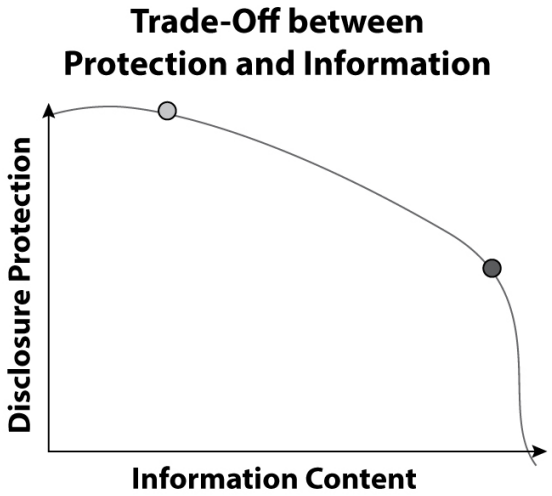
> "Beneficial technology sometimes has unintended consequences. Sometimes products or services that make life simpler or more convenient also put us at greater risk. As more devices monitor and track our lives, they also gather copious amounts of personal and sensitive data that could be compromised or exposed. *The potential attack surface is exponentially greater when almost everything you touch is somehow collecting data or gathering information about you*…. (Emphasis added)
>
> Consider the vehicles being produced today. Most have GPS navigation of some sort, and some take it a step further with active monitoring systems. The car knows where you've been, where you are, what direction you're going, and how fast you're driving at any given moment. That's great if you get in an accident and a service is able to dispatch emergency response automatically. But what happens if the police start tapping that data to issue speeding tickets, or a divorce attorney can subpoena your vehicle's location data to prove where you were at a given point in time?"

Cautions and concerns about "creepiness" like those noted above could jeopardize development and availability of societal benefits that could otherwise be made possible by new improvements in technology.

[11] See http://www.cisco.com/c/en/us/solutions/collateral/enterprise/cisco-on-cisco/Cisco_IT_Trends_IoE_Is_the_New_Economy.html
[12] See https://blogs.rsa.com/creepy-factor-internet-things/

5. Growing Tensions Between Data Privacy and Data Value

"No matter what the arena – finance, health care, or national security – questions surrounding the provision of personal data are always the same: how much benefit vs. how much risk? *Who handles this data, and can those individuals be trusted? How do organizations guard against data misuse."*[13] (Emphasis added)



The above quotes and graph[14] (which plots on the opposing x and y axis "Disclosure Protection" and value of "Information Content") highlight the escalating tension between the objectives of data privacy and data value. These tensions are quickly reaching a "boiling point" due to increasing volume, velocity and variety (the '3Vs') of big data and associated data computational capabilities; together these factors increase the daily likelihood that any given data element alone, or in combination with other data elements, can be linked to a specific individual and / or that individual's traits or habits. These factors generate anxiety and serious concerns, causing state, federal and international data privacy laws, rules and regulations to be invoked in order to protect the privacy rights of Data Subjects.[15]

[13] Footnote 10, *supra*.

[14] Barth-Jones, Daniel, *Statistical De-identification: Challenges and Solutions*, HHS Workshop on the HIPAA Privacy Rule's De-Identification Standard at http://hhshipaaprivacy.com/assets/5/resources/Panel2_Barth-Jones.pdf

[15] There are more than 400 regulations worldwide mandating information security and data protection requirements making it difficult for organizations to comply. If a data element alone, or in combination with other data elements, can be linked or is linkable to a specific individual, it may constitute personal data protected under International, federal, and / or state statutes, rules and regulations (e.g., in the European Union, Article 8 of the European Convention on Human Rights; in Canada, the Personal Information Protection and Electronic Documents Act (PIPEDA); in the U.S., the Health Insurance Portability and Accountability Act (HIPAA), the Fair Credit Reporting Act (FCRA), and the Family Educational Rights and Privacy Act (FERPA); in California, the California Online Privacy Protection Act (OPPA)).

Regardless of differing interpretations of what constitutes PII,[16] potential adverse consumer reaction to perceived or real vulnerabilities jeopardizes market acceptance of initiatives like genomic research, personalized medicine and the Internet of Things (IoT) / Internet of Everything (IoE). Furthermore, meeting the demands of both the public and regulators diminishes data value and imposes new compliance costs on parties in various data processing streams.[17]

McKinsey & Company article entitled *Views From The Front Lines Of The Data-Analytics Revolution*[18] reported that data-analytics leaders "…were unanimous in their view that placing more control of information in the hands of consumers, along with building their trust, is the right path forward." FTC Commissioner Julie Brill quoted from this same McKinsey report during a speech entitled *The Internet of Things: Building Trust and Maximizing Benefits Through Consumer Control* when she noted that "Privacy has become the third rail in the public discussion of big data."[19] Consider, then, if every

[16] Resolution of appropriate definitions for, and treatment of, terms such as "personally identifiable information," "personal information" or "PII" is beyond the scope of this letter; similarly, a discussion of policy issues pertaining to whether Data Subjects should directly manage data and / or have Trusted Parties manage data on their behalf is beyond the scope of this letter. Anonos Dynamic Anonymity provides technology tools to ensure controls are available for data use in accordance with applicable regulations. While policy tools by themselves provide clarity as to when situations involve wrongdoing or inappropriate use of data, policy-based remedies by themselves may be "too little, too late" if a Data Subject suffers identity theft, loss of credit, denial of time sensitive services, discrimination, etc. An analogy exists between the need for technology tools as a compliment to policy tools and the need for injunctive relief in appropriate circumstance as a compliment to legal remedies. (See http://www.academia.edu/1548128/The_Inadequacy_of_Damages_as_a_Remedy_for_Breach_of_Contract). Without the benefit of complimentary technology tools, there may be "no adequate remedy by policy alone." In addition, the ability of Anonos Dynamic Anonymity to address tensions between current and future data use practices and Fair Information Practice Principles (FIPPs) (see http://www.nist.gov/nstic/NSTIC-FIPPs.pdf) by providing users with direct control over the use of their data may help reconcile differences between European Union "fundamental right" and United States "balancing of right to freedom of expression / commerce" perspectives on data privacy protection.

[17] For example, effective as of September 22, 2014 under the HIPAA / HITECH final rule, business associates are now directly liable for HIPAA compliance with civil and potential criminal penalties for noncompliance. *Because business associates were not directly and primarily liable under HIPAA before the final rule, many service providers to covered entities may not realize they are now HIPAA regulated business associates.* Three categories of service providers are specifically identified as business associates under the final rule: (1) health information organizations, e-prescribing gateways, and other people or entities that provide data transmission services to a covered entity with respect to protected health information and that require access on a routine basis to such protected health information; (2) people or entities that offer personal health records to one or more individuals on behalf of a covered entity; and (3*) subcontractors that create, receive, maintain or transmit protected health information on behalf of business associates. The addition of subcontractors means that all requirements and obligations that apply to business associates of a covered entity also apply to all downstream service providers.* For example, a data storage company that has access to protected health information is a business associate (regardless of whether they view the information) and both data transmission services and personal health record vendors may be business associates based on facts and circumstances - i.e., if the vendor has access to protected health information in order to perform its duties and responsibilities (regardless of whether it actually exercises this access) the vendor is a business associate.

[18] See http://www.mckinsey.com/Insights/Business_Technology/Views_from_the_front_lines_of_the_data_analytics_revolution?cid=other-eml-alt-mkq-mck-oth-1403

[19] See http://www.ftc.gov/system/files/documents/public_statements/289531/140314fordhamprivacyspeech.pdf

attribute of any device or entity in the IoT / IoE could be de-identified in such a way that it maximized both privacy and value while at the same time preserving the trust of Data Subjects. *We believe this is achievable by using Anonos dynamic de-identifiers (DDIDs) as highlighted in the Detailed Example of One Potential Architectural Structure for Technical Implementation of DDIDs for Data Capture attached on page 28 as Appendix A.*

6. The Need for Trusted Alternative Approaches that Provide Control

Data Subjects, so as not to respond negatively in the face of big data capabilities, must trust that their data is private, protected and used for intended purposes, all while maintaining maximum value. Current frameworks, such as static anonymity, neither build trust with Data Subjects nor effectively serve businesses, researchers, or government. The proliferation of technology, while opening some doors, has pitted privacy interests against those of economic growth and national security.

In order to achieve health, education and scientific advances – and to produce promising commercial applications – traditional approaches to data capture, transmission, storage and analysis must be revisited. New alternatives are necessary to meet the demands of the public and regulators while simultaneously maintaining high data value.

Dynamic Data Obscurity provides significant improvements over static approaches to protecting privacy. Dynamic Anonymity and Anonos-enabled Circles of Trust (CoT) combine Dynamic Data Obscurity together with well-established data capture, transmission, storage and analysis techniques as well as with Privacy Enhancing Technologies (PETs). This combination provides optimal data privacy protection while still preserving high value data analysis capabilities.

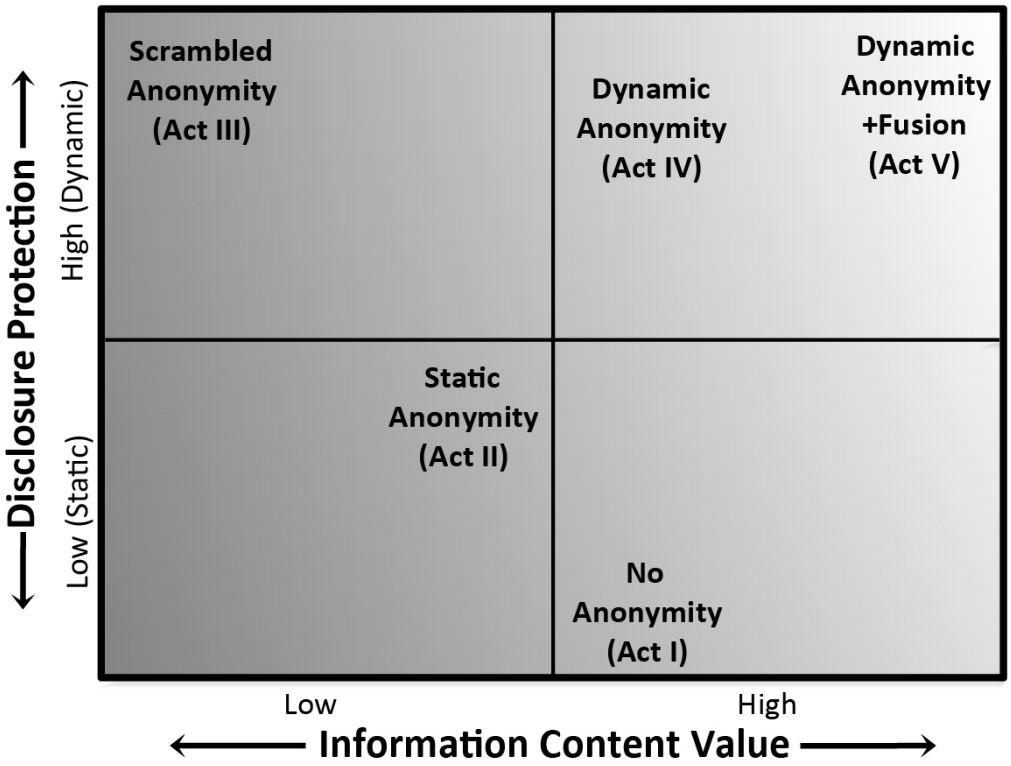
Dynamic Anonymity enables, through the use of dynamic de-identifiers (DDIDs), commercial opportunities like the Internet of Things (IoT) / Internet of Everything (IoE) to achieve their enormous market potential by enabling Data Subjects to employ a trusted approach to keeping data private and protected, and to ensuring that data is used only for intended purposes – *thereby avoiding potential negative consumer reactions*. At the same time, Dynamic Anonymity minimizes or eliminates many of the concerns of governmental organizations charged with protecting the rights of Data Subjects. In this way, parties can save money and conduct better research while *minimizing the cost of complying with data privacy regulations.*

7. Dynamic Data Obscurity Supported by Dynamic Anonymity / Circles of Trust (CoT) based on DDIDs

The use of Dynamic Anonymity and Anonos-enabled Circles of Trust (CoT) enhances privacy, protects personal data and increases the value of that personal data by leveraging DDIDs to enable data to be collected and managed by trusted third parties ("Trusted Parties") in accordance with permissions established by, or on behalf of, Data Subjects.

The concept of Dynamic Anonymity is a key element in the Anonos solution. Attached as Appendix B on page 35 is a five act "play" illustrating the differences between non-existent (Act I), static (Act II),

and scrambled anonymity (Act III), on the one hand; and Dynamic Anonymity with Anonos (Act IV), and the benefits of Dynamic Anonymity and big data fusion[20] (Act V), on the other hand. Figure 1 below highlights the fact that only with the latter two are both privacy and value of data maximized.



NOTES: No Anonymity (Act I) does not provide maximum value – due to overload of information – and provides no protection. Static Anonymity (Act II) provides medium privacy and medium value. Scrambled Anonymity (Act III) reduces value to near zero while providing strong privacy. Highest value and strongest protection comes from Dynamic Anonymity (Act IV) and Dynamic Anonymity plus Fusion (Act V).

Figure 1

[20] The PCAST Report states that "data fusion occurs when data from different sources are brought into contact and new facts emerge." See http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf

Dynamic Anonymity / Circles of Trust (CoT)

- *Dynamic Anonymity is premised on the principle that static anonymity is an illusion and that the use of static identifiers is fundamentally flawed.* The Anonos patent-pending system[21] dynamically segments and applies re-assignable dynamic de-identifiers (DDIDs) to data stream elements at various stages,[22] minimizing the risk of information being unintentionally shared in transit, in use or at rest, while maintaining the ability of Trusted Parties – and of no others – to re-stitch the data stream elements.

- Cleartext primary keys may be used internally within a Circle of Trust to identify Data Subjects; however, these keys may not be shared outside the Circle of Trust. Rather, Dynamic Anonymity uses dynamically changing and re-assignable compound keys outside of a Circle of Trust - each comprised of (i) a DDID and (ii) the time period / purpose for which the DDID is associated with a given Data Subject. Information regarding this association never need be made available outside of the Circle of Trust (nor may it be reconstructable, since the connections between a Data Subject and data pertaining to them need not contain any recoverable information leading back to the Data Subject – i.e., the connections may be inherently incomputable).

- Dynamic Anonymity enhances privacy and personal data protection capabilities in distributed platforms / fragmented ecosystems while providing superior access to, and use of, data in accordance with policies established by, or on behalf of, Data Subjects. In this manner, everyone – including those who elect to use either closed or distributed systems – benefits from enhanced data privacy and anonymity.

- Dynamic Anonymity delivers certain immediate benefits without modification to existing business and technology practices. With the use of dynamically changeable, temporally unique, and re-assignable DDIDs, current systems and processes (e.g., web browsers and data analytic engines) would not recognize relationships between and among data elements. These systems and processes can process information using existing capabilities without creating inferences, correlations, profiles or conclusions except as expressly authorized by Data Subjects and Trusted Parties via a Circle of Trust (CoT). However, additional significant benefits would arise from new business and technology practices that leverage specific attributes and capabilities of DDIDs and Dynamic Anonymity.

---

[21] U.S. Patent Applications No. 13/764,773; 61/675,815; 61/832,087; 61/899,096; 61/938,631; 61/941,242; 61/944,565; 61/945,821; 61/948,575; 61/969,194; 61/974,442; 61/988,373; 61/ 992,441; 61/994,076; 61/994,715; 61/994,721; 62/001,127; 14/298,723; 62/015,431; 62/019,987; 62/037,703; 62/043,238; 62/045,321; 62/051,270; 62/055,669; 62/059,882; 14/529,960; 14/530,304; 14/530,339 and International PCT Patent Applications No. PCT US13/52159 and PCT/US14/63520. Anonos, Dynamic Anonymity, CoT, Dynamic De-Identifier, and DDID are trademarks of Anonos Inc.

[22] While dynamic segmentation may include time lapse, it is more likely determined by activity, location and / or subject matter.

Figure 2 below represents the concept of an Anonos-enabled Circle of Trust (CoT) from a Trusted Party perspective. *Note first that the Data Subject is included on the diagram at the bottom left.* Diagrams of most current data use systems do not include Data Subjects since participation by Data Subjects generally takes the form of a binary decision whether to agree to "take-it-or-leave-it" online terms and conditions using the traditional "notice and consent" model.[23] After that initial point, the Data Subject typically loses all power to affect what happens to their data since "they are the product, not the customer."[24] It is well acknowledged that this is a broken model for the digital age and provides few effective limitations on current or future use of data.

An Anonos-enabled Circle of Trust (CoT) leverages Dynamic Anonymity to empower a Data Subject to whom data pertains (a "Subject User") to select from pre-set policies (similar to, but also easily more granular than, selecting a low, medium or high level of protection when installing anti-virus software) that translate into discrete dynamic permissions. Alternatively, a Subject User may select a "Custom" option to specify more detailed dynamic parameters (similar to selecting custom installation options when installing application software).

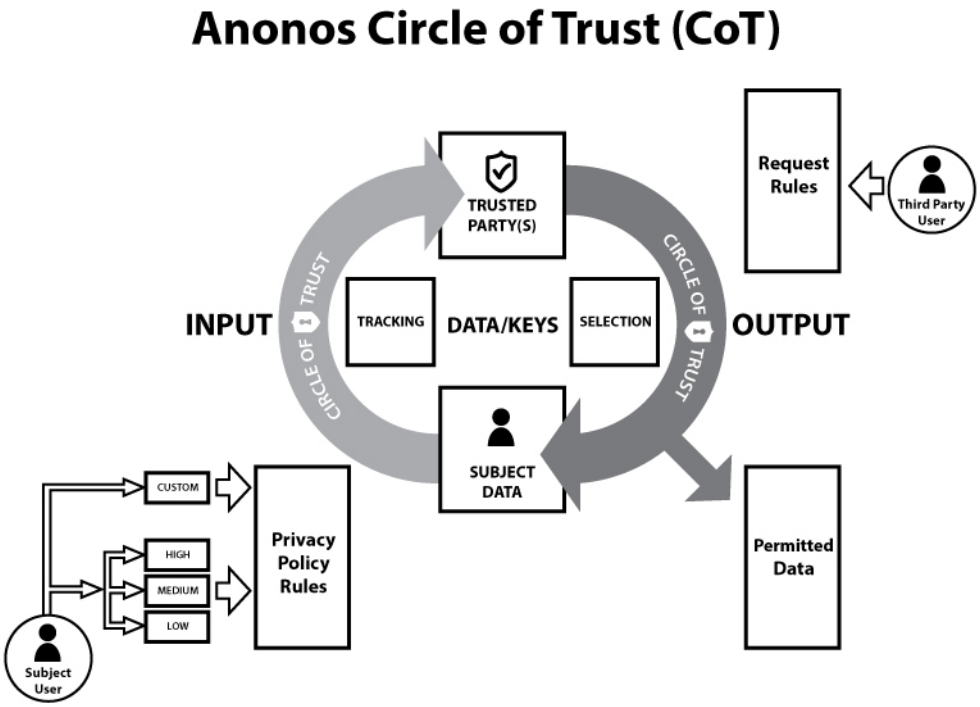


Figure 2

[23] Take-it-leave-it "notice and consent" online terms and conditions are acknowledged as a "market failure" in the PCAST Report.

[24] Bruce Schneier, security expert and author, said in a 2010 speech at the RSA Europe security conference in London, "Don't make the mistake of thinking you're Facebook's customer, you're not – you're the product….Its customers are the advertisers." See http://www.information-age.com/technology/security/1290603/facebook-is-%22deliberately-killing-privacy%22-says-schneier

Privacy Policy Rules relate to allowable operations such as what data can be used by whom, for what purpose, what time period, etc. Rules may also specify desired anonymization levels such as when / where / how to use dynamic de-identifiers (DDIDs) for dynamic obscuring (as more fully described herein) in the context of providing anonymity for the identity and / or activities of a Data Subject, when to use other Privacy Enhancing Techniques (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 whom data pertains (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" in Figure 2 represents data available for sharing with parties external to the Circle of Trust (CoT) that satisfies Privacy Policy Rules established by Subject Users and / or Request Rules established by Third Party Users.

It should be noted that there may be more than one Trusted Party working cooperatively in connection with a single Anonos-enabled Circle of Trust and that Data Subjects may be participants in any number of Circles of Trust. Circles of Trust can be implemented by means of a centralized or federated model for increased security. Arrows in Figure 2 represent data movement; data inputs and outputs will contain different information.

Figure 3 below represents the concept of an Anonos-enabled Circle of Trust (CoT) from a Data Subject perspective.

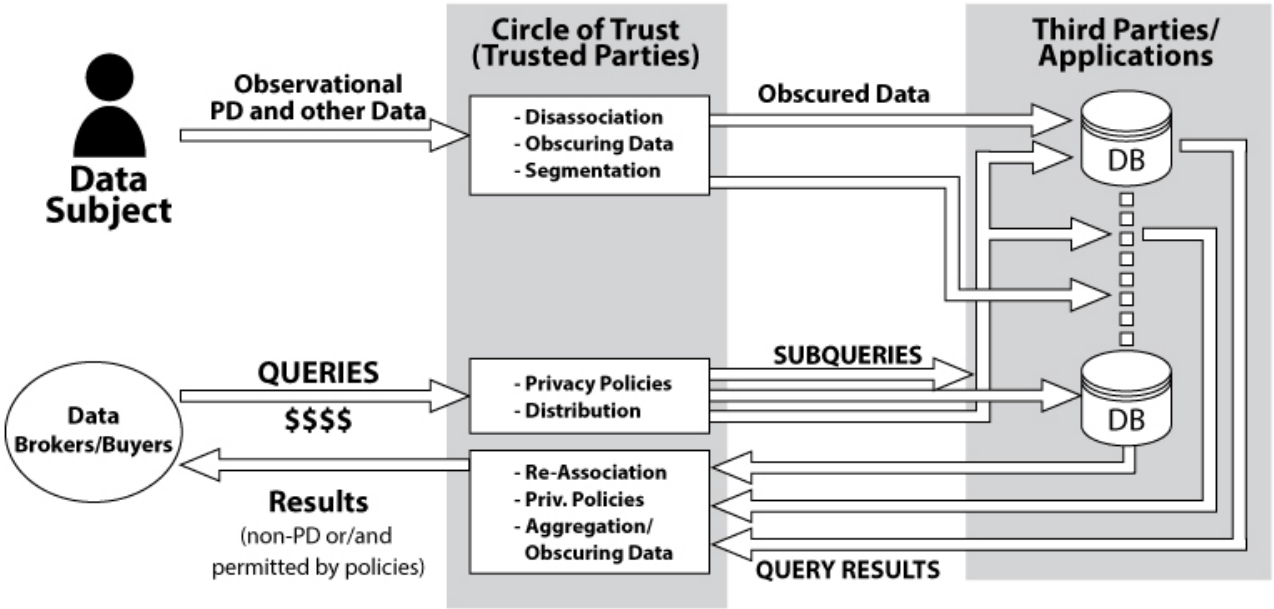


Figure 3

8. Challenges of Data Cleansing and Further Benefits of Dynamic Data Obscurity

Anonos-enabled Dynamic Data Obscurity provides benefits at four distinct points of data processing:

A. Data Capture;

B. Data Transmission / Storage;

C. Data Analysis; and

D. Privacy / Anonymity Control.

At each point data is protected in accordance with privacy policies and / or rules specified by, or on behalf of, Data Subject(s) to whom that data pertains. This will be explained in progressively greater detail below, and illustrated via several examples.

8.A. Data Capture

In applications where a static identifier would typically be associated with capture of data pertaining to a Data Subject, Anonos-enabled Dynamic Dat Obscurity can provide:

1. A dynamic de-identifier (or DDID) that may change over time (triggered by a lapse of time, change in purpose, temporary cessation in activity, or change in virtual or physical location) limiting the ability to track, profile or otherwise associate observational data with a Data Subject.

2. An association from each DDID to the applicable Data Subject, stored and known only within the applicable Circle of Trust (CoT).

3. Anonos-enabled Dynamic Data Obscurity also offers the optional ability to store data associated with DDIDs within a CoT.

A key feature of Dynamic Anonymity is the ability to anonymize and segregate data elements at the data element level rather than at the data record level – i.e., at the level of data elements representing actions, activities, processes and / or traits of a Data Subject rather than at the level of the Data Subject. Anonos-enabled Circles of Trust retain relationship information between and among data elements and Data Subjects to permit reassembly according to privacy policies and / or rules established by, and / or on behalf of, Data Subjects.

Example: Search Engine

Consider a person who frequently uses a particular search engine. Currently, the search engine assigns the person (via their browser) a "cookie" or other digital footprint tracker that persists for months or years, against which an ever-increasing stream of observational data (e.g. search terms, links clicked, location data) is then accumulated and, very likely, analyzed and further aggregated by multiple parties – often revealing personally identifiable information without knowing consent by the Data Subject.

Anonos-enabled Dynamic Data Obscurity can leverage the natural response of a search engine to create a new cookie / digital footprint tracker for each Data Subject perceived to be interacting with the search engine for the first time. Clearing history, cache, cookie / digital footprint tracker, and associated data will cause the search engine to generate a new cookie / digital footprint tracker for the Data Subject. An Anonos-enabled Circle of Trust (CoT) can store information pertaining to associations of cookies / digital footprint trackers to the Data Subject, and optionally also store a list of queries and selected links.

With this approach, the search engine would still have access to aggregate data – trending search terms, popular websites, ad clicks, etc. – but would be prevented from drawing inferences related to the Data Subject based on observational data. If / as authorized by privacy policies and / or rules established by, and / or on behalf of, the Data Subject, the CoT could enable the search engine to perform more detailed analysis. This could be implemented using an HTTP proxy or browser extension, requiring no modification to (or cooperation from) an existing search engine.

In the past, anonymous tracking cookies were supposed to have solved the problem of how to support both privacy and analytics. However, anonymous tracking cookies failed to achieve this goal because all the data was housed together and associated with random static identifiers that made it too easy to generate Personal Data (PD), thereby nullifying or attenuating the value of the static "anonymous" identifiers.[25] Anonos Dynamic Anonymity overcomes these shortcomings by employing dynamically changing and re-assignable DDIDs, storing the resulting DDID associations and obscuring keys within Anonos-enabled Circles of Trust, and providing a unique interaction model enabling participation between and among Data Subjects and Trusted Parties / third-party participants.

As highlighted on page 39 in the ACT III Curtain of the "play" attached as Appendix B, incognito mode private browsing and similar attempts at "scrambled anonymity" provide a level of anonymity for Data Subjects but at the cost of the loss of personalized offerings made possible by ongoing advances in technology (e.g., the Internet of Things (IoT), personalized medicine, etc.) and potential loss of social accountability. While these approaches may make use of dynamically changing identifiers, the value of information is largely destroyed. In contrast, with no anonymity, the value of information is harmed by the mere quantity of it, reducing the Return-on-investment (ROI) on any given piece of information, which is why the popular belief that more gross data equals greater value is not actually true.

8.B. Data Transmission / Storage

An Anonos-enabled CoT is composed of one or more Trusted Parties, each of which may offer one or more independent data storage facilities, as well as secure means (via Anonos Dynamic Anonymity or Privacy Enhancing Technology (PET)-enabled means of obfuscation and / or encryption) to segment and transmit sensitive data to these stores.

[25] See Footnote 1 and Footnotes 3 through 5, *supra*.

Alternatively, Anonos-compliant application developers could choose to only store the Data Subject-to-DDID associations within the CoT, and instead to use Anonos Dynamic Anonymity-defined procedures to obscure, encrypt, and / or segment data (or utilize Anonos-enabled toolkits for such procedures); allowing applications to safely store generated or collected information in their own facilities, without loss of context or business value.

Example: Networked Blood Pressure Monitor

For example, imagine a smartphone application that can track both geolocation and blood pressure levels. Using Anonos-enabled Dynamic Data Obscurity, such a device could split data into two streams, each obscured such that either stream, if intercepted and / or compromised (or even examined once stored), would not reveal Personal Data (PD) without the addition of critical information protected within the CoT.

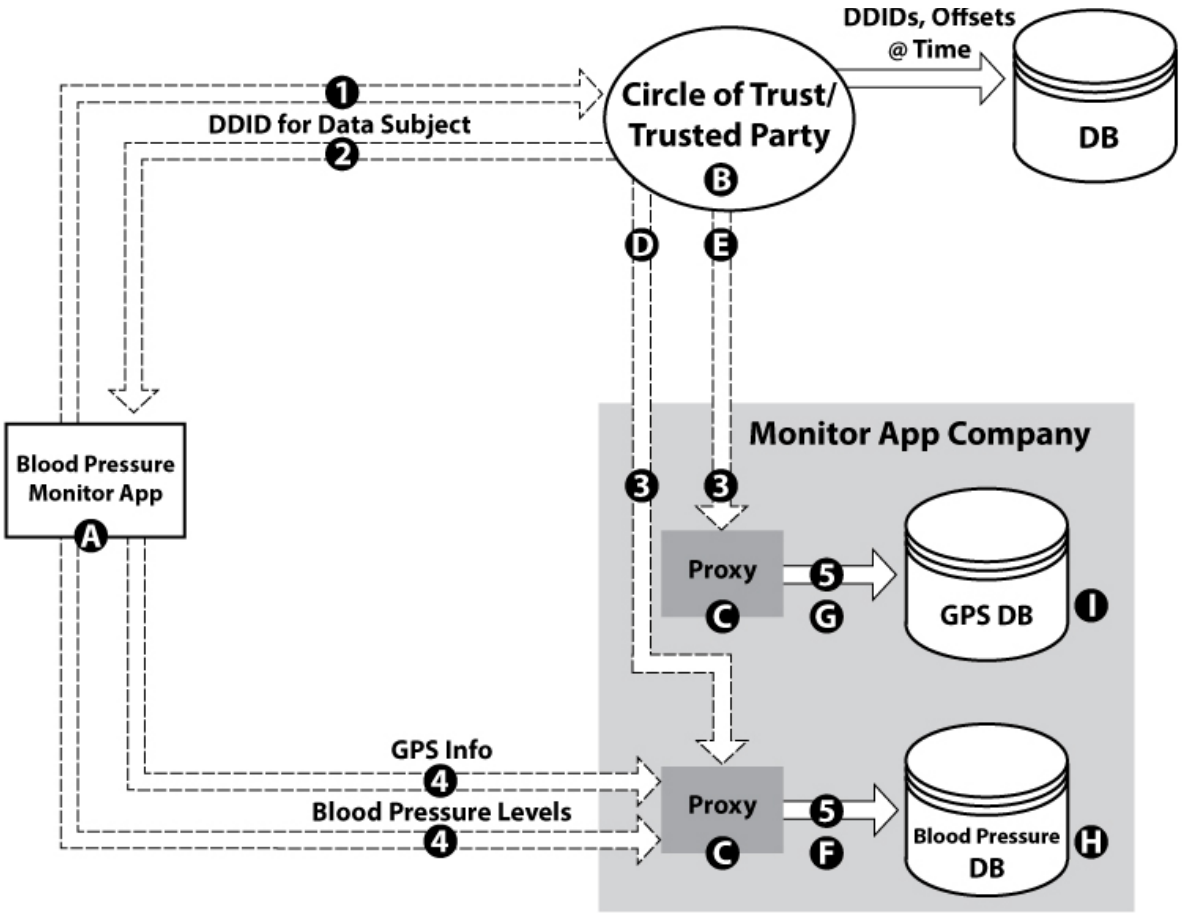


Figure 4

Figure 4 illustrates:

1. The blood pressure monitoring application (A) contacts a Trusted Party within a Circle of Trust (B) requesting a DDID for the Data Subject patient.

2. The CoT Trusted Party provides a DDID for the Data Subject.

3. An application operated by the Trusted Party sends back two sets of periodically-changing information (one for GPS data, one for blood pressure levels), each consisting of DDIDs, offsets (to obscure blood pressure level data and geographic position), and encryption keys; refreshed for each new time period. (These are also stored to a database for later use.)

4. The monitor application transmits two encrypted and obscured streams of data to an Anonos-controlled "proxy" application or network appliance (C) within its corporate network. (Here, both location and levels have a periodically changing offset applied to them.)

5. The "proxy" (C) uses the streams of data (D & E) from the Trusted Party (containing only decryption keys) to convert the transmitted data into "plaintext." The proxy also hides the incoming IP address and provides stream(s) (containing multiple Data Subjects' information) of DDIDs and obscured blood pressure level data (F) or GPS locations (G) to the corresponding databases (H) and (I).

At each point outside the Circle of Trust (and outside the smartphone itself) the patient's data is protected; no Personal Data (PD) is made available or ever produced.

- Transmissions to and from the Trusted Party (1, 2) have no privacy-harming Personal Data, nor is any stored in the Trusted Party's database.

- Location and blood pressure levels (4) are transmitted separately (intercepting any one stream reveals nothing), keyed by DDIDs, and obscured so that even the data itself neither reveals nor contains anything, directly or indirectly, about the patient's true location or blood pressure levels.

- The Anonos proxies (C) must be connected to the Trusted Party in order to decrypt the data (preventing a man-in-the-middle attack). Each merges multiple streams of data together, after decryption, so that the originating IP address cannot be associated with its decrypted data.

- Once at rest, when residing in two separate databases (H and I), the blood pressure levels and location data each have different sets of DDIDs, so that even the hosting company cannot draw any association between the two, much less link each set of data to the Data Subject who produced it.

As noted previously, similar techniques to those employed in this example have been employed in the past:

- Segmenting data;
- Encrypting and obfuscating data during transmission; and
- Employing distribution, obfuscation and security during storage.

However, Anonos-enabled Dynamic Data Obscurity leverages and improves upon these approaches by:

- *Employing dynamically changing and re-assignable DDIDs to obscure data at the data element (versus data record) level;*
- *Storing resulting DDID associations / obscuring keys within Anonos-enabled Circles of Trust; and*
- *Providing a unique interaction model for enabling participation between and among Data Subjects and Trusted Parties / third-party participants.*

8.C. Data Analysis

Traditional techniques for data "cleansing" (also referred to as data cleaning and data scrubbing) paradoxically suffer from two different and antithetical kinds of problems.

1. A given data cleansing technique can simply be ineffective. Despite earnest efforts, or even use of legally sanctioned techniques to obscure Personal Data[26], it may be still possible to identify the Data Subjects and Personal Data from "cleansed" data. Three famous examples:

    a. In the mid-1990s, the Massachusetts Group Insurance Commission (GIC) released data on individual hospital visits by state employees in order to aid important research. Latanya Sweeney, then an MIT graduate student, purchased the Cambridge voter-registration records, and by linking the two data sets, which individually were completely innocuous, she was able to re-identify then-Massachusetts Governor Bill Weld's GIC entry despite the fact that it had been "anonymized," with all obvious identifiers, such as name, address, and Social Security number, removed.

    b. In 2006, Arvind Narayanan, then a graduate student at UT-Austin, together with his advisor, showed that by linking the "anonymized" Netflix dataset to the Internet Movie Database

[26] For example, the American Medical Informatics Association states at http://jamia.bmj.com/content/20/1/29.full that "HIPAA sets forth methodologies for de-identifying health data; once such data are de-identified, they are no longer subject to HIPAA regulations and can be used for any purpose. Concerns have been raised about the sufficiency of HIPAA de-identification methodologies, the lack of legal accountability for unauthorized re-identification of de-identified data, and insufficient public transparency about de-identified data uses."

(IMDb), in which viewers review movies, often under their own names, many Netflix users could be re-identified.

c. In 2013, a team led by Dr. Yaniv Erlich, of the Whitehead Institute for Biomedical Research, re-identified men who had participated in the 1000 Genomes Project – an international consortium to place, in an open online database, the sequenced genomes of (as it turns out, 2500) "unidentified" people – who had also participated in a study of Mormon families in Utah.

2. More effective data cleansing techniques may reduce the business value of that data – that is, many obfuscation techniques are lossy.

The Anonos approach to data privacy provides a way to avoid both pitfalls, simultaneously.

Example: Serving Patients With Sexually Transmitted Diseases (STD)

To illustrate, consider the task of choosing a location for a new clinic to serve patients who are 20 to 30 years old with sexually transmitted diseases (STDs). One "cleansed"[27] data set may show the incidence of STDs, aggregated by neighborhood to protect privacy. Another data set may show how many patients reside in each neighborhood. But, even when these are aggregated, one cannot know exactly how many identified cases of STDs fall into particular age ranges.

Anonos alleviates this dilemma by supporting two different modes of analysis.

In cases where data must be exposed externally (that is, outside the CoT), Personal Data elements can be obscured or encoded as DDIDs, with the resulting associations stored inside the CoT. Additionally, when required, the data (or field) type identifiers can also be obscured in a similar manner.

Later, after analysis is performed, the results of that analysis can then (when permitted) be associated back with the original Data Subjects, field types, and values.

Another way Anonos enables lossless analysis is through the use of federated, anonymized queries, either among different Trusted Parties within a CoT, different data stores within the same Trusted Party, or between Trusted Parties and application developers whose data stores reside outside the CoT.

Consider again the problem of choosing where to site a clinic to serve patients who are between 20 and 30 years old with STDs. The Anonos system improves upon existing techniques by allowing the target query to span multiple data stores and dividing it up such that each participant does not know what purpose it serves, so there is no risk of divulging PD.

[27] In order to protect patient privacy, HIPAA requires Personal Data that reveals Personal Health Information / (PHI) to be "cleansed" to limit disclosure of PHI; this limits availability of Personal Data to support personalized medicine / medical research.

In this scenario, the query for the number of patients who are 20 – 30 years old with STDs within a set of (sufficiently large) geographic areas is presented to numerous Trusted Parties within the Anonos Circle of Trust. This aggregate query is then broken down into several steps, such as:

1. Find patients between 20 – 30 years of age in some broad geographic area.
2. Select only those with STDs.
3. Select only those whose privacy policies allow this level of analysis.
4. "Join" those results to the home addresses of those patients.
5. Aggregate these results by neighborhood, revealing only counts of patients.

   The actions needed to satisfy this query could span completely different data stores, in different organizations – nonetheless protected and facilitated by the Circle of Trust.

For Example:

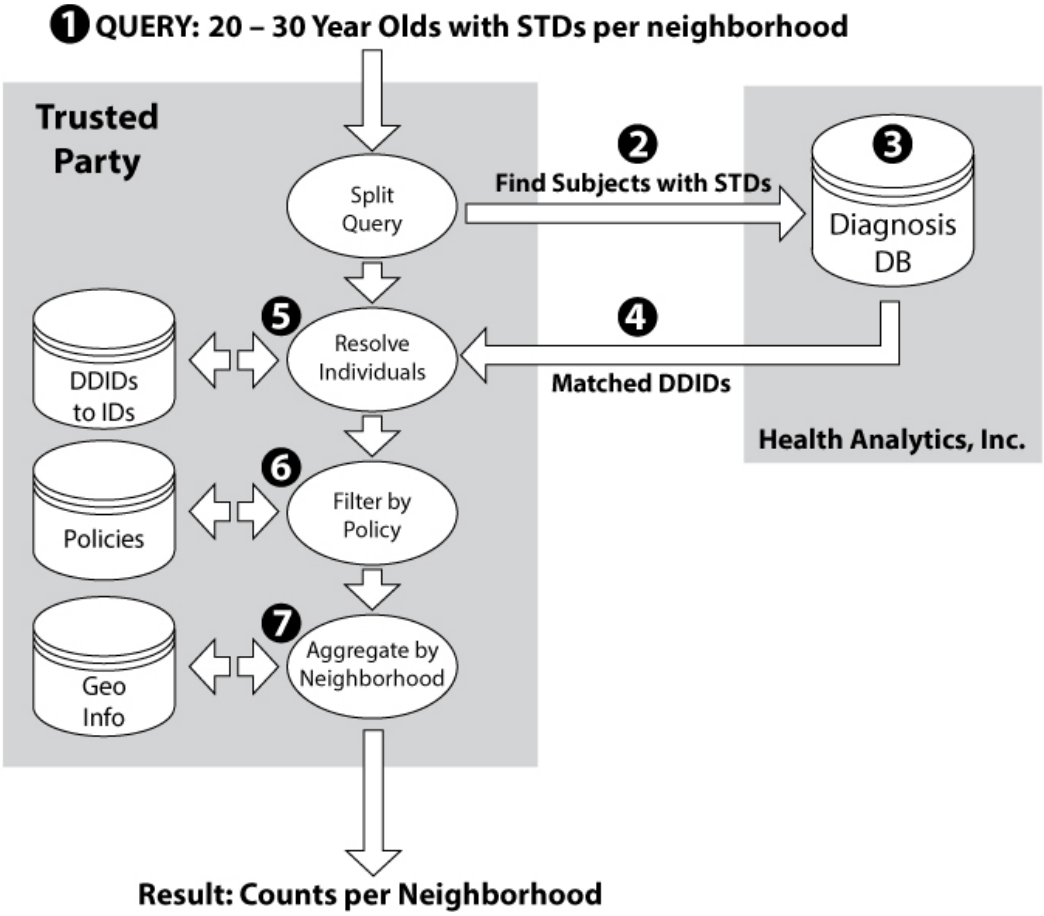


Figure 5

1. The prospective clinic owners send a query to a Trusted Party, asking to find individuals who are between 20 – 30 years old with STDs.

2. The Trusted Party contacts healthcare-related data stores to find individuals who are between 20 – 30 years old with STDs.

3. The healthcare-related data stores (which store diagnoses by DDIDs rather than by identifiable keys) find matching records.

4. Matching DDIDs are then transmitted back to the Trusted Party.

5. The Trusted Party then resolves these DDIDs to unveil identified individuals.

6. The Trusted Party filters that list by those whose privacy policies allow this particular kind of query.

7. The CoT then uses a database of their addresses to aggregate counts (or incidence frequency, if the query is incomplete) by neighborhood, producing the desired result.

In this scenario, companies operating healthcare-related databases do not need to know (or divulge) the identity, location, or other potentially identifiable information of the patients whose data they possess. The records they possess are keyed by DDID, and also potentially obscured, so that no Personal Data is generated when performing the specified query, nor when transmitting results.

Note that the party posing the query does not have access to this information. Their only interaction with the CoT consists of posing a question and receiving a high-level, aggregated, non-PD result. Note that not having access to this information in no way affects the quality, accuracy or precision of the end result. Anonos thus eliminates Personal Data that contributes nothing to the end result and that only serves to weaken privacy without any attendant benefit to any other party. By filtering out irrelevant data, the analysis of which would otherwise consume time and resources, this process actually *increases* the utility and value of the information received.

Personal Data is only produced temporarily, *within the Circle of Trust* managed by the Trusted Party (the appropriate place for such information) — such as when the DDIDs are resolved. Such operations are *transient* and leave no lasting trace other than the intended query result, and could also be confined to certain dedicated servers for increased security. The use of DDIDs in the context of Anonos-enabled Circles of Trust avoids potential shortcomings of normal data analytics that could generate discriminatory or even identifiable results.

8.D. Privacy / Anonymity Control

In order to protect Personal Data, Anonos-enabled Dynamic Data Obscurity may employ a multiple means of measuring, specifying, and enforcing data privacy / anonymity:

1. A system for determining a privacy / anonymity level for each potential kind of exposure for data associated with the subject. These privacy / anonymity levels may consist of a continuum of

discrete values (between the extremes of complete privacy / anonymity and complete public exposure), and / or a mathematical specification of such (an "Anonymity Measure Score" or "AMS").

2. Rules: actions allowed or limited by policies regarding data. (For example: "share," "update.")

3. Polices, which associate access levels, permissions and data with each other, thus granting or denying certain levels of access to data on the basis of one or more criteria, including data type, time, organization seeking access, etc.

A Data Subject's privacy policy rules may also be combined with, or limited by, statutory policies. (For example, medical data in the US must be protected in accordance with HIPAA.)

Additionally, if allowed by the Trusted Party and with the data owner's consent, offers to modify or grant specific and limited permissions may be presented to, and accepted by, Data Subjects.

Anonos-enabled Dynamic Data Obscurity also improves upon existing frameworks by using privacy / anonymity level determinations to prevent inappropriate use of data, which is obscured and only analyzed, whether from inside or outside a Circle of Trust, in a manner consistent with each Data Subject's specified privacy level.

Example: Offering a Coupon

A shoe manufacturer wishes to send a coupon for a new line of shoes to people who have recently performed web searches related to the sport of running within a certain city. In exchange for offering discounts on the shoes, the manufacturer wishes to receive qualified consumers' email and / or home addresses, and to send those who redeem the coupon a survey to assess their satisfaction with the new shoe.

Such an interaction might look like this:

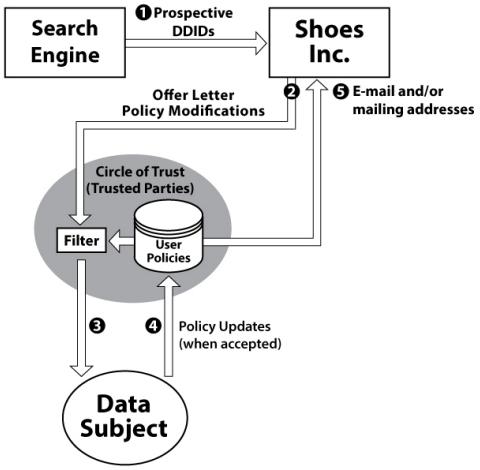


Figure 6

Explanation:

1. The manufacturer, outside the CoT, purchases a list of matching DDIDs from a search engine.

2. The DDIDs are submitted to one or more Trusted Parties, accompanied by an offer letter and a policy modification allowing access (upon acceptance) to Data Subjects' email and / or home addresses.

3. Each Trusted Party then forwards the offer letter to the Data Subjects matching those DDIDs (provided they have opted-in to receiving such an offer).

4. If a Data Subject recipient accepts the offer, the recipient's policy is updated with (perhaps temporally-limited) permission for exposing their home and/or e-mail addresses to the shoe company.

5. The shoe manufacturer, now part of the CoT, but only with respect to this specific offer and only in the most limited sense, then receives a list of e-mail and home addresses of those who wish to receive the coupons. Note that this list is necessarily highly targeted and accurate and therefore of maximum value to the shoe manufacturer. This is precisely how the Anonos CoT, by increasing privacy, also increases value. The shoe manufacturer may be assured that all mailings done this way will be sent to those with substantial interest in the manufacturers' offer.

Note that sending coupons to qualified prospects / customers by means of Dynamic Anonymity and Anonos-enabled Circles of Trust could help avoid negative public reaction to big data algorithms like the one exposed in the 2012 *The New York Times* article entitled *How Companies Learn Your Secrets*.[28] This article reported how a big data algorithm determined that a teenage girl was pregnant by analyzing her purchasing history resulting in Target sending coupons for pregnancy items to the girl at her family home thereby notifying her father of the pregnancy before she told him. Worldwide public reaction to this story ranged from a scathing law review article on tensions between potential benefits of big data and resulting privacy harms[29] to demands for new laws and regulations.[30]

Example: A Physician Views Blood Pressure Levels

Recall the earlier example where a GPS-enabled blood pressure monitor securely stored patients' locations and blood pressure levels via Anonos-enabled Dynamic Data Obscurity. Anonos-enabled Dynamic Data Obscurity could be leveraged to:

[28] See http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html
[29] See Crawford, Kate and Schultz, Jason, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2325784
[30] See article entitled *Why Big Data Has Made Your Privacy a Thing of the Past*, available at http://www.theguardian.com/technology/2013/oct/06/big-data-predictive-analytics-privacy

1. Avoid imposition of HIPAA data handling obligations on business associates[31] involved in data processing flows if data in their possession does not constitute Personal Data (PD).

2. Ensure that access to, and use of the data, by the physician satisfies HIPAA obligations.

   Note that the following scenario assumes that both a Data Subject patient and his / her physician have accounts inside the Circle of Trust.

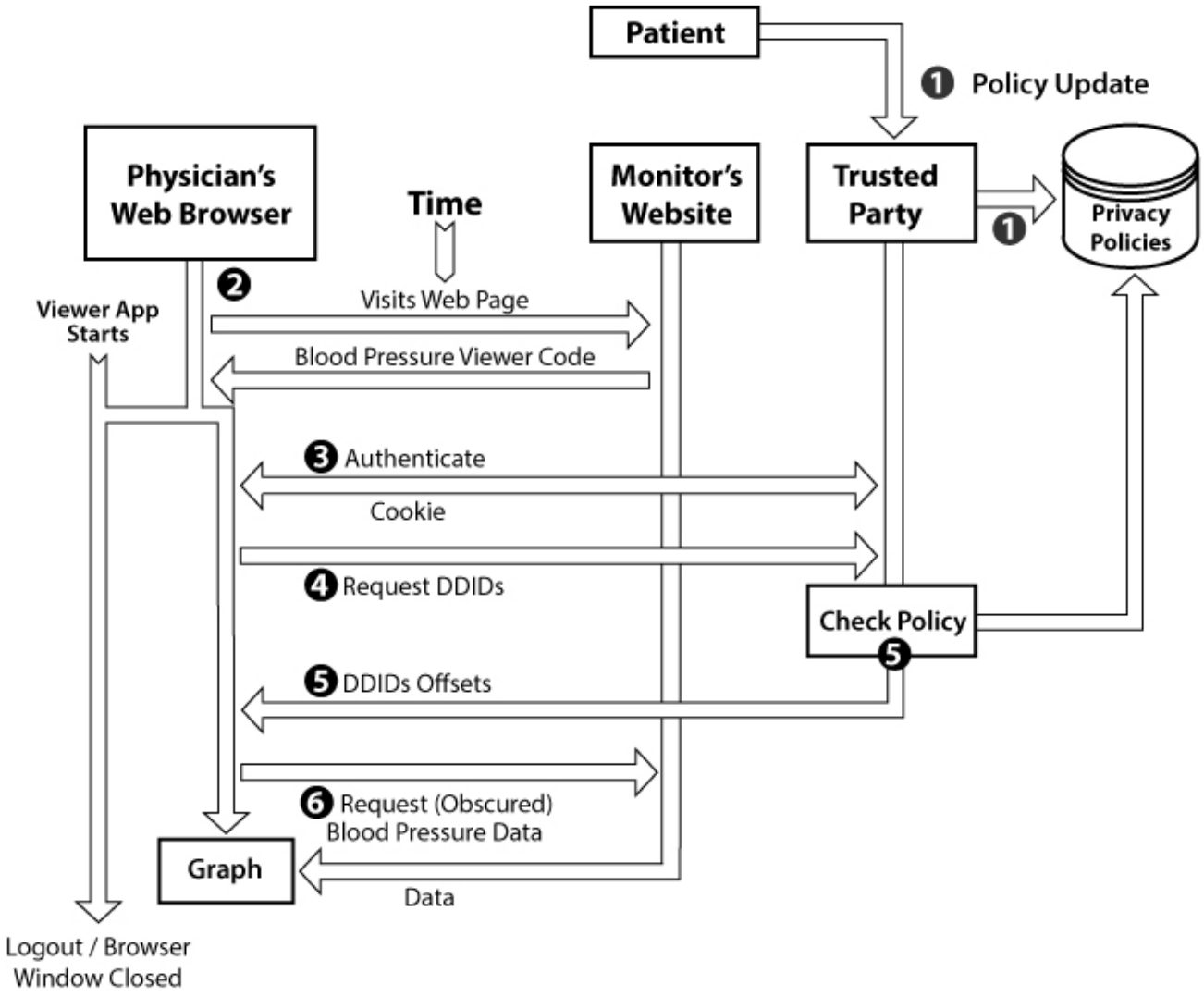


Figure 7

[31] See Footnote 17, *supra*.

Explanation:

1. The monitoring application cooperates with the patient's Trusted Party to allow the patient to update his / her privacy policy rules so that his / her physician can now access his / her blood pressure levels (but not his / her GPS location data). Note that this grant can be temporary (analogous to the temporally limited nature of photographs that can be shared with Snapchat – the grant expires after a period of time) – or ongoing.

2. The physician (via his / her web browser) browses to the blood pressure monitor's web site, which launches a JavaScript-based blood pressure level viewer application which thus *runs in the physician's browser*, and not on the monitor company's servers (i.e., that the stitching together of data necessary to make it personally identifiable is done via the Trusted Party server which is itself trusted – see steps 4 and 5 below).

3. The blood pressure-level viewing application asks the physician to log in via her Trusted Party (similar to the way many applications allow you to authenticate using a Facebook or Google account), and receives a session cookie that continues to identify them to that party.

4. After the physician selects a range of time to view, the viewer application requests the relevant DDIDs and offsets from the Trusted Party, for that patient.

5. The Trusted Party validates the physician's access to this information (checking the patient's privacy policy rules) and then returns the DDIDs and offsets.

6. The viewer application then contacts its own corporate website, requests the blood pressure data corresponding to those DDIDs, receives the result, applies the offsets, and renders the blood pressure levels as a graph.

At this point, the image on the physician's screen is HIPAA-protected PHI data. If the physician prints the data, that paper will be subject to HIPAA. When the physician is done viewing the graph, he / she logs out or closes the browser, the application ends, and the data is erased.

Note that re-identified HIPAA-controlled data only resides in the physician's browser. The original blood pressure level data stored in the application provider's databases remains untouched and obscured. The Trusted Party's data remains unaffected as well.

Also note that the permission to view the blood pressure data is enforced within the Circle of Trust. It is not enforced (as is common practice today) merely by the viewer application – or only by the application's backend servers. This means that an adversary could not gain unauthorized access to the data merely by hacking into the blood pressure level viewer application, because the data would not be there in any usable or identifiable form. The dynamic data obscuring capabilities of Dynamic Anonymity DDIDs combined with the dynamic data privacy control capabilities of an Anonos-enabled "Circle of Trust," maximize both data privacy and value to support personalized medicine / medical research.

Anonos appreciates the opportunity to submit this letter to the U.S. Federal Trade Commission and to the Republic of Mauritius Data Protection Office.

Respectfully Submitted,

M. Gary LaFever
Co-Founder

Ted Myerson
Co-Founder

cc: Julie Brill, Commissioner – U.S. Federal Trade Commission

Mme Axelle Lemaire, Minister of State for the Digital Sector – attached to French Minister of the Economy, Industry and the Digital Sector

Alexander Macgillivray, Deputy Chief Technology Officer – U.S. Office of Science and Technology Policy

John Morris, Associate Administrator and Director of Internet Policy – U.S. National Telecommunications & Information Association

Günther Oettinger, European Digital Economy and Society Commissioner-designate – European Commission

Megan Smith, Chief Technology Officer – U.S. Office of Science and Technology Policy

Ashkan Soltani, Chief Technologist – U.S. Federal Trade Commission

Appendix A

Detailed Example of One Potential Architectural Structure for Technical Implementation of DDIDs for Data Capture

Dynamic De-Identifiers (DDIDs)

A dynamic de-identifier DDID is a temporally-bounded pseudonym which both refers to and obscures the value of (i) a primary key referencing a Data Subject, (ii) the value of an attribute of that Subject (e.g. a ZIP code), and/or (iii) the *kind or type* of data being associated with the Subject (e.g. the *fact* that some encoded value was a ZIP code).

DDIDs protect data because there is no discernable, inherent, nor computable relationship between their content and the values (cleartext) to which they refer. Additionally, the association between a given DDID and its cleartext value is not exposed outside the Circle of Trust (CoT). Unlike static identifiers, an obscured value or key need not have the same associated DDID when used in a different context, for a different purpose, or at a different time.

DDIDs can be either generated within the Circle of Trust, or if the above criteria are satisfied, external IDs can be used as DDIDs (e.g. cookies issued by the search engine described in Section 8.A. on page 15).

DDIDs are Time-Bounded

As mentioned, DDID associations are *temporally-bounded*, by which we mean that, even within the same context, and with regard to a single type of data (e.g. ZIP code), a particular DDID may refer to one value at one time, but may (if desired) also refer to another value at a different time.

This necessarily implies that in order to decode or expose the meaning of a particular DDID, an application must also retain knowledge of the time to which that DDID applied.

This knowledge may be *explicit* – that is, the assignment time may also be part of the record or document in which the DDID was stored – or it may be *implicit* – for example, an entire data set may have been obscured as a batch, and presumed (regardless of how long processing actually takes) to have occupied the same instant – and thus have only one consistent set of DDID mappings per field type. In order to reconstitute such data, one would also need to supply some reference to the corresponding set of DDID / value associations (stored within the CoT).

DDIDs are Purpose-Bounded

Note that DDIDs are also bounded by *context* or *purpose* – meaning the same DDID can recur in multiple contexts, even at the same time. For example, consider a stream of records, each of which

contain a Social Security Number (SSN) and ZIP code, and which all occupy a single time block. In such a case, a particular DDID may be used both as a replacement for a ZIP code, and also as a replacement for an SSN.

As above, this implies that some indication of that context (e.g. was this a ZIP code or SSN?) will be necessary to obtain the cleartext to which that DDID referred.

Replacing Data with DDIDs

Consider the task of replacing a single stream of data – the same kind of data (e.g. ZIP codes or SSNs), occupying the same time block – with DDIDs. A (Java-like) "pseudocode" description of an Application Programming Interface (API) that carries out such behavior might look like this:

```
interface DDIDMap {
   DDID protect(Value cleartext);
   Value expose(DDID ddid);
}
```

In English, "interface" means that we're defining a collection of functions (named "DDIDMap") that operate on the same underlying data. Data types are here denoted with initial upper-case letters (e.g. "DDID"), and variable or function parameter names are denoted with initial lower-case letters (e.g. the "cleartext" function parameter must be data of type "Value" – where "Value" is just a stand-in for any kind of data which can be obscured: IDs, quantities, names, ZIP codes, etc.).

One function, "protect()", accepts some cleartext value and returns a corresponding DDID. If that value has been seen previously, its previously-assigned DDID will be returned. If it has not been encountered before, a new DDID (so-far unique to this data set) will be generated, associated with that value, and then returned.

The other function, "expose()", reverses this process: when a DDID is passed to it, it looks up and returns the cleartext value, which was previously encoded as that DDID. If the given DDID has never been seen before, it fails with an indication of error.

The data managed by these operations, then, is a two-way mapping from each cleartext value to the DDID that replaced it, and from the DDID back to the original value.

Note that although we've said that a given DDID can only refer to a single value, it **is** possible, if desired, to implement a variant version of this algorithm that allows a value to be associated with *more than one* DDID.

Managing DDID Maps by Time and Purpose

Recall that the above bidirectional DDID-to-value map operates (i) upon a single kind of data (that is, having the same type, context, and purpose), and (ii) within the same time block. In order to support operations across multiple times and contexts, we can posit another API which gives us the an appropriate DDID-to-value map for a given time and purpose:

```
interface DDIDMapManager {
   DDIDMap getMap(Context context, Time time);
}
```

Here, "context" is (or emits) a key that refers to a particular kind of data being obscured. (In other Anonos documents, this is sometimes also called the "association key" or "A_K".) For example, the context might be the name of the table and column in which data to be obscured will reside (e.g. "employee.salary"). It could also include other non-other chronological indications of purpose or scope.

The "time" parameter indicates the instant at which the DDID is being (or was) associated with its cleartext value. Since DDID-to-value maps span a *block* of time, and there are many time instances within a block, this implies there exists some function (used internally, within this API, thus not shown above) that finds the time block associated which each given time. (More on this in a moment.)

DDID Generation and Time-Blocking Strategies

Note that different kinds of data can employ different DDID replacement strategies. In addition to those mentioned in the next two sections, DDIDs can vary in size, whether they're universally unique or just unique to that data set (or time block), what kind of encoding they use (e.g., integers or text), etc. And although DDID generation should typically be random, one might also wish to employ deterministic or pseudo-random DDID generators for demonstration, testing, or debugging purposes.

Unique or Reused DDIDs

One potential strategy allows a particular DDID to be assigned to **two different** Data Subjects in the same context, but during two different time blocks. For example, within the same collection of time-anchored records, the DDID "X3Q" might at one moment (in one time block) refer to (for example) "80228", and later (in another time block), "12124". (We'll call this strategy "DDID reuse.")

An alternative is to disallow such "reuse" – and stipulate that a given DDID, in the same context, can only refer to a single Subject. (Although the subject may still receive different DDIDs over time.)

The choice between these two strategies involves a tradeoff between increased obscurity and the ease with which one may perform aggregation queries on obscured data.

Imagine we wish to count patients per postal code. If postal codes DDIDs are unique, we can aggregate counts per DDID, and then ask the CoT to finish the query by resolving those DDIDs to their

corresponding postal codes, and aggregating again. But if we have "reused" DDIDs, then we must send the entire list of DDIDs and corresponding times to the CoT for resolution (and aggregation) – because we can't be sure that two instances of the same DDID refer to the same value.

DDID Time Blocks

Implementations also have freedom to choose different strategies for segmenting DDID maps by time. Blocks of time may vary by size and / or time offset; sizes can be fixed, random, or determined by number of records assigned per time. (Note that employing an infinite-sized time block (for a given context) gives behavior equivalent to using "static" identifiers.)

Implementation

Although there may be many strategies for creating new DDIDs, the API for generating such DDIDs can look (essentially) identical, regardless of which strategy is implemented "under the hood".

For example:

```
interface DDIDFactory {
   DDID createDDID();
}
```

Next, consider the task of determining what time block was associated with a given DDID assignment. Since a time block can contain many instances of time, we'll need some kind of a "time key" (abbreviated "T_K" in other Anonos documents) to each time block. This implies the need for a function to obtain the appropriate key for any time instant:

```
TimeKey timeKey = getTimeKey(Time time);
```

Further, note that both time-blocking and DDID-generation strategies depend upon the *kind* of data which are being obscured. In short, they are both associated with a given "context" (which includes or implies a notion of data type and usage), meaning that the "Context" API must offer at least one function supporting each:

```
interface Context {
   TimeKey getTimeKey(Time time);
   DDIDFactory createDDIDFactory();
}
```

Given these two additional functions, we can imagine that the implementation of "getMap()" in "DDIDManager" (shown previously) might look something like this:

```
DDIDMap getMap(Context context, Time time) {
   TimeKey timeKey = context.getTimeKey(time);
   DDIDMap map = getExistingMap(context, timeKey);
```

```
   if (map was not found) then
      DDIDFactory factory = context.createDDIDFactory();
      map = createMap(factory);
      storeNewMap(context, timeKey, map);
   endif
   return map;
}
```

(Here, "getExistingMap()" is some function that finds the map assigned to the given context and time key, "createMap()" creates a map which will use the given DDID factory, and "storeNewMap()" associates a newly-created map with the context and time key by which it will be retrieved later.)

Using Context to Obscure Data and Attribute Types

Recall again that the Anonos platform defines three kinds of data which can be protected: (i) primary keys which refer to subjects (e.g. employee ID), (ii) attribute data associated with, but not unique to, a Subject (e.g. employee postal code), and (iii) the indication of a disassociated (obscured) data element's *type*, itself (an "association key", or "A_K").

Each of these can be achieved by defining a different context: first we'll discuss (i) and (ii), which are both achieved by obscuring data values (replacing them with "replacement key" DDIDs, abbreviated as "R_K" elsewhere). We will address (iii) the indication of a disassociated (obscured) data element's *type*, below.

Consider a trivial example: an order table recording which customers bought products on a given day. Each record has a day number, a customer ID, and a product ID. We want to obscure this data for use or analysis by some third party, who is outside the CoT. In particular, we wish to obscure the customer and product IDs, but leave the day numbers intact.

To do so, we could create two "Context" instances: one for "Customer ID", and one for "Product ID". Although DDIDs, should ideally be random, for our purposes, let's assume that our "DDIDFactory" will create integer DDIDs sequentially, starting from 0. Further, assume that each DDID map spans only three days, so after three days, a new set of DDID mappings will be used. This also implies that DDIDs will be "reused" – the same DDID can refer to different values when used different blocks. (This is not an ideal encoding strategy and is used here only for illustration purposes.)

Here is some cleartext sample data:

Day	Customer ID	Product ID
1	500	ZZZ
2	600	XXX
3	600	YYY
4	700	TTT
5	500	YYY

6	600	TTT

After being obscured (as specified above), this data would become:

Day	Customer ID	Product ID
1	0	0
2	1	1
3	1	2
4	0	0
5	1	1
6	2	1

To understand this, you read down each column, and think in groups of three days (the first time block of DDIDs covers, for each obscured field, days 1-3, and the second covers 4-6).

For the first three days, customer ID is: 500, 600, 600
The resulting encoding is: 0, 1, 1 (note that 600 is repeated, so its DDID, 1, is also repeated.)

For the second three days, customer ID is: 700, 600, 500
And (starting over from 0), the result is: 0, 1, 2 (note that 500 was 0 before, now it's 2)

Product ID uses a separate context, and thus stream of DDIDs, so it also starts from zero:
For the first time block (XXX, YYY, TTT) becomes (0, 1, 2)
For the second time block (TTT, YYY, TTT) becomes (0, 1, 0)

Another "Context" could be employed to obscure the indication of a disassociated (obscured) data element's *type* (iii above), where the column names are examples of Attribute Keys (A_K)). This could be done using one DDID-to-value mapping for the whole set (effectively substituting DDID for the column names), or in time blocks (as with the other fields in this example) such that (if an appropriately random DDID generation strategy were employed) the affected records could not be analyzed without the assistance of the Circle of Trust.

Notes on Locality and Time

The example APIs defined above presume that when data is encoded, the encoding time is passed with each datum or record. This is only necessary when DDIDs are being "reused" within the same context (and thus time is needed to discriminate between the two potential meanings of that DDID). When a DDID is only assigned to one value per context, that DDID is sufficient to discover the (single) original value.

Time could also become an issue where "reused" DDIDs are being employed across different systems, which might have slightly different notions of time. If it is not possible to pass the time associated with a DDID encoding, a (chronological) "buffer" could be employed to prevent a DDID from being re-used

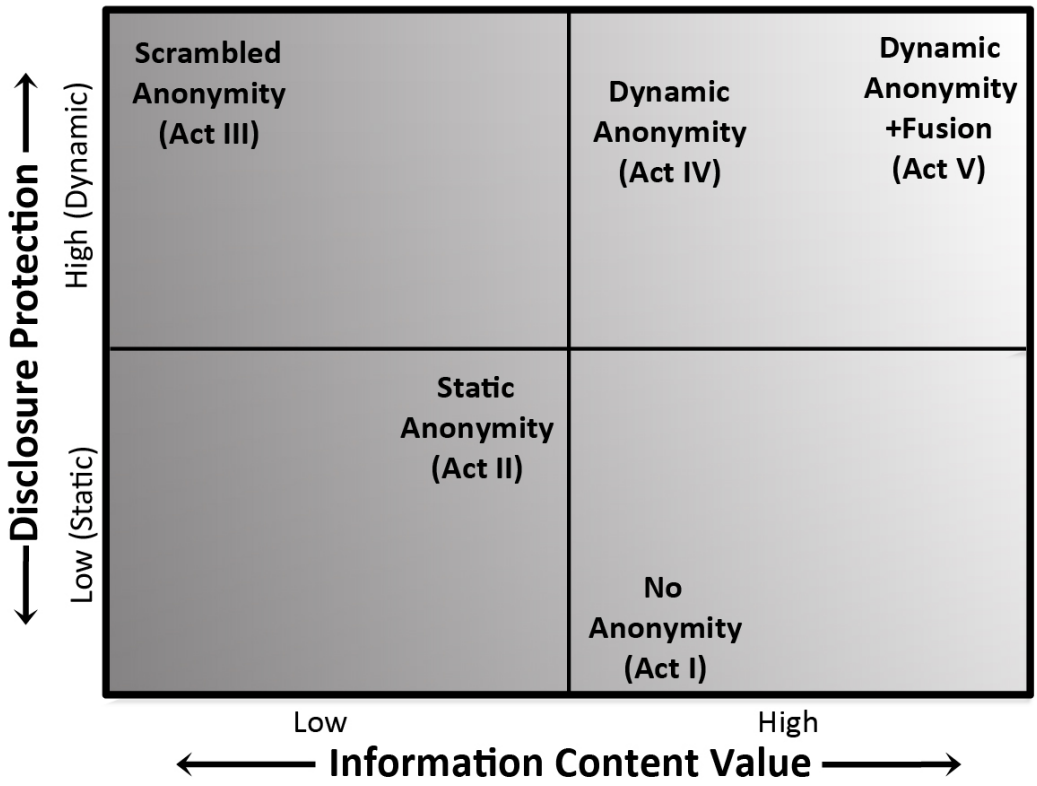
too close to it's original assignment. And when it is possible to pass the time associated with the data to be encoded, the time could be "sanity-checked" against the local system clock: skew within a small window (smaller than the DDID reuse buffer) could be tolerated, whereas larger differences would trigger an error report.

Finally, note that there is also flexibility regarding *where* data is being encoded: data could be streamed to a machine residing within the CoT, and then sent along to its destination after encoding. But, alternatively, the *encoding* portions of the above algorithms could be run *outside* the Circle of Trust, provided that the resulting DDID-to-value associations were (a) not stored on the local host, and (b) safely (e.g. using encryption, and with appropriate safeguards against data loss) streamed to a CoT host for persistence, lowering latency in critical applications.

Appendix B

Five Act "Play" Illustrating Anonos Dynamic Anonymity

DISCLAIMER – This "play" is not intended as an example of the Anonos invention as a whole – it is intended as a simple example of differences between non-existent, static, and scrambled anonymity, on the one hand; and Dynamic Anonymity (with Anonos), and the benefits of large scale Dynamic Anonymity with fusion, on the other hand, with the addition of offline information being input into a system. While Anonos could be implemented in a similar manner, this "play" only illustrates a partial application of Dynamic Anonymity.



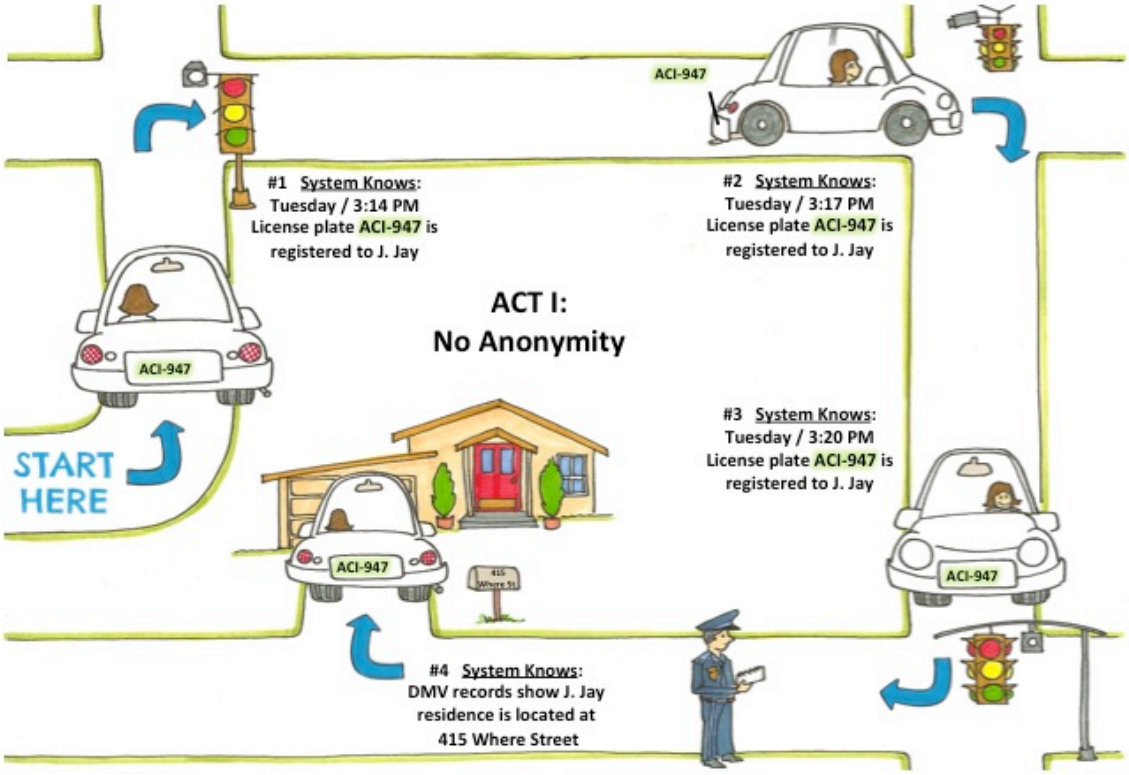
NOTES: No Anonymity (Act I) does not provide maximum value – due to overload of information – and provides no protection. Static Anonymity (Act II) provides medium privacy and medium value. Scrambled Anonymity (Act III) reduces value to near zero while providing strong privacy. Highest value and strongest protection comes from Dynamic Anonymity (Act IV) and Dynamic Anonymity plus Fusion (Act V).

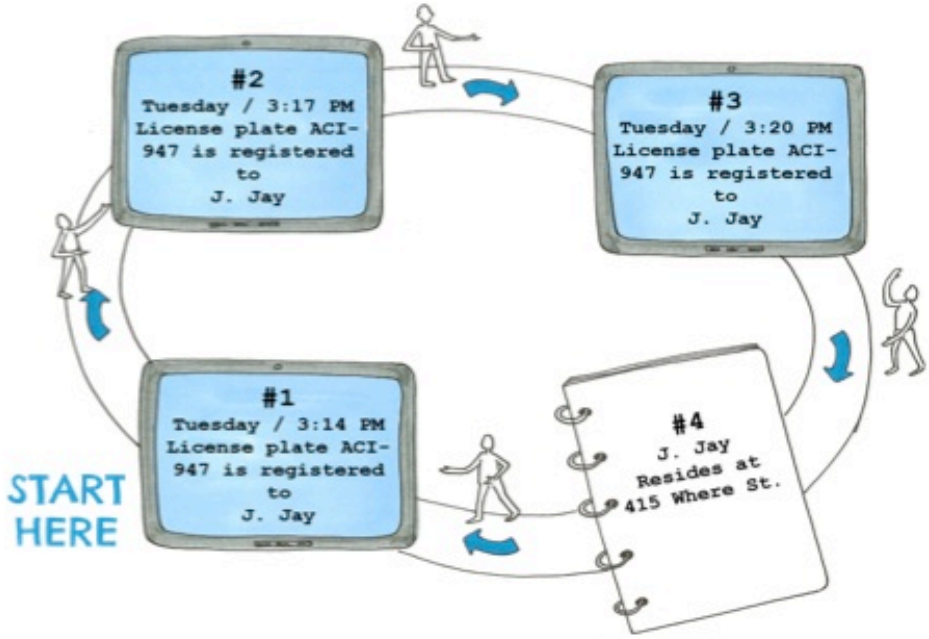
The following "play" represents a hypothetical, pseudo real-world parallel of how Anonos Dynamic Anonymity is unique compared to other anonymization practices and how it supports big data ***fusion.****

- *Imagine a car drives through three traffic lights, each of which has an automatic license plate reading (ALPR) camera. The car then turns a corner and parks in a garage located at 415 Where Street.*

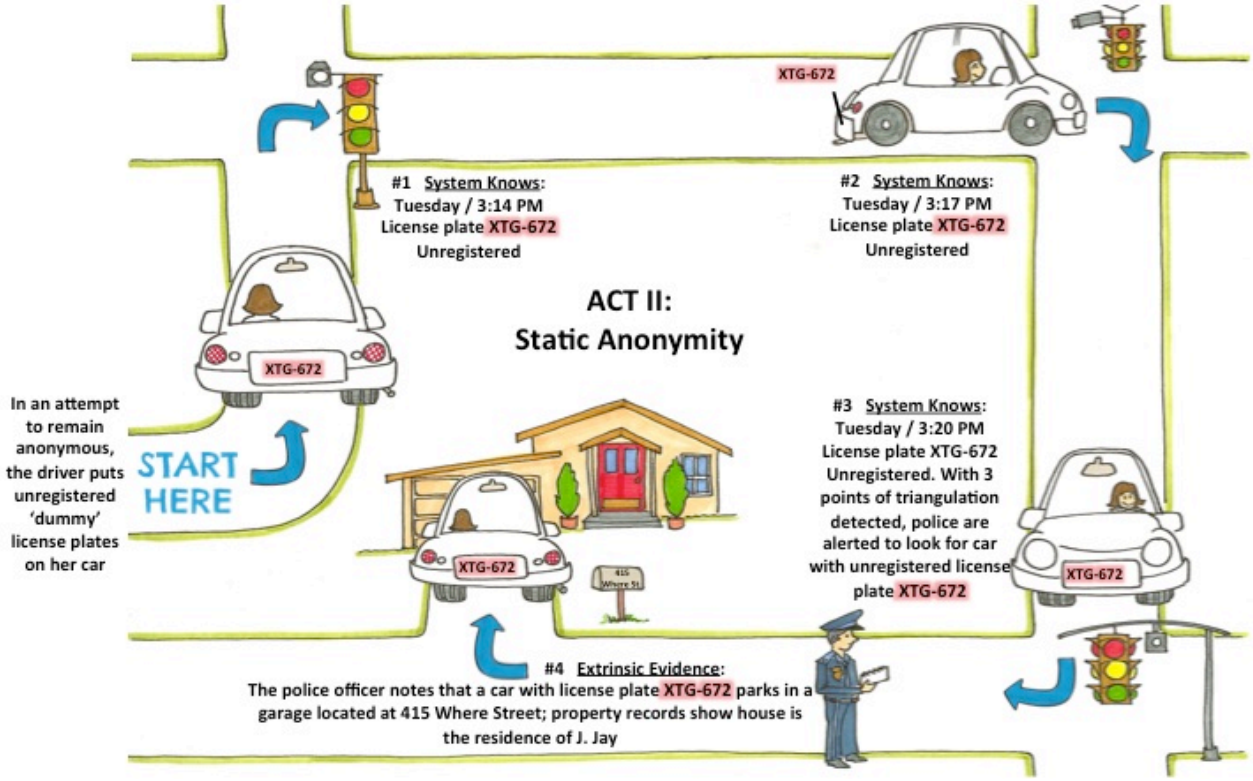
- *In this play, the female driver represents a data subject and the police officer represents a corporate / governmental authority figure.*

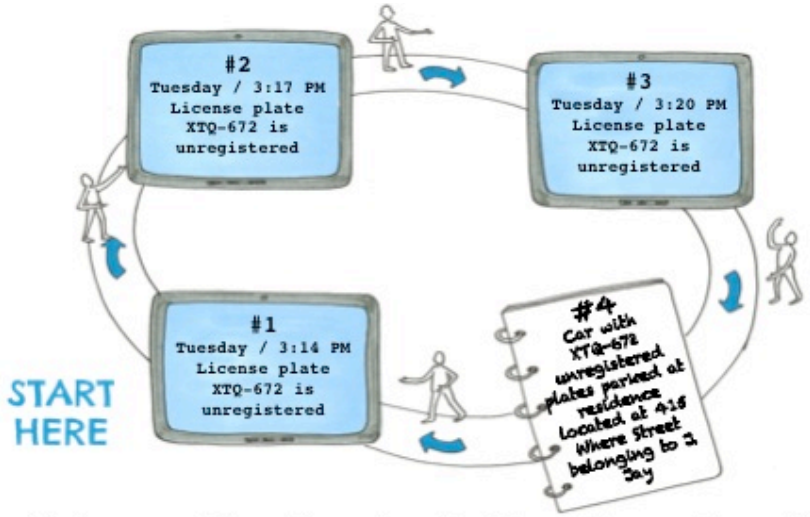
* The May 2014 U.S. President's Council of Advisors on Science and Technology (PCAST) report entitled *Big Data and Privacy: A Technological Perspective Data* states that "data fusion occurs when data from different sources are brought into contact and new facts emerge." See http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf



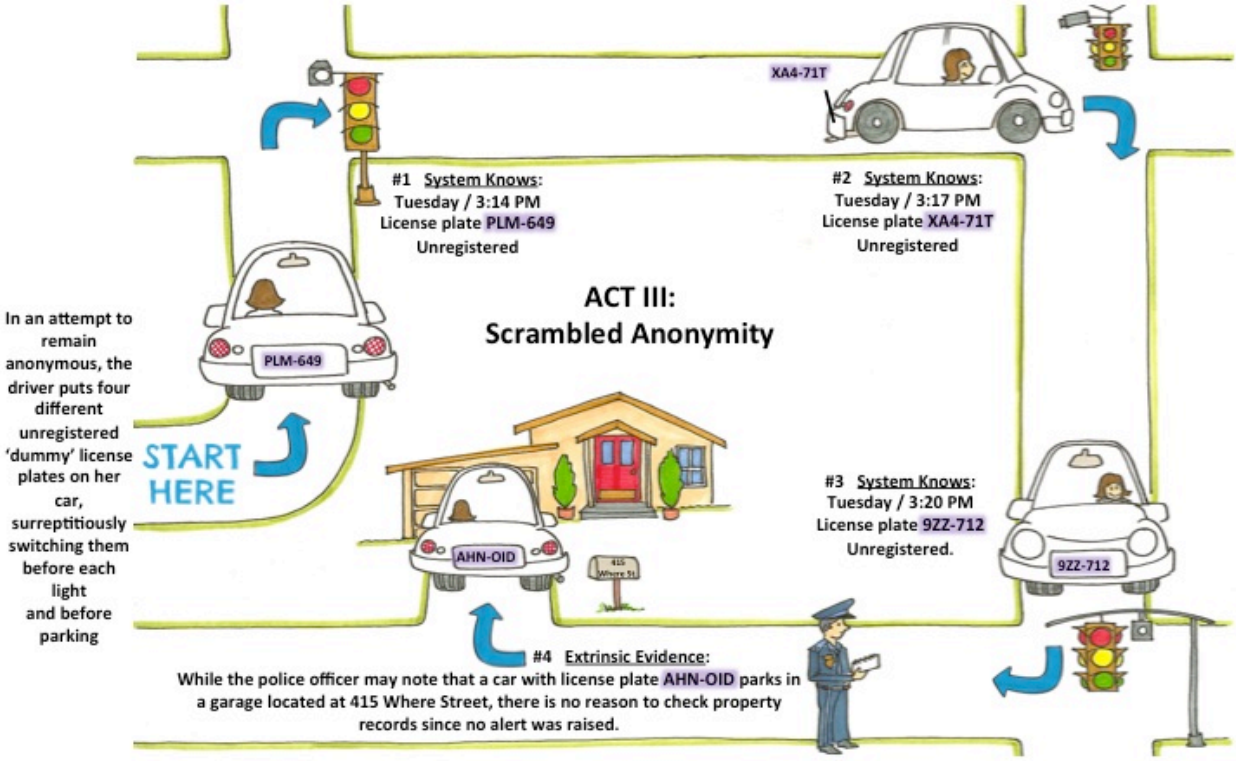


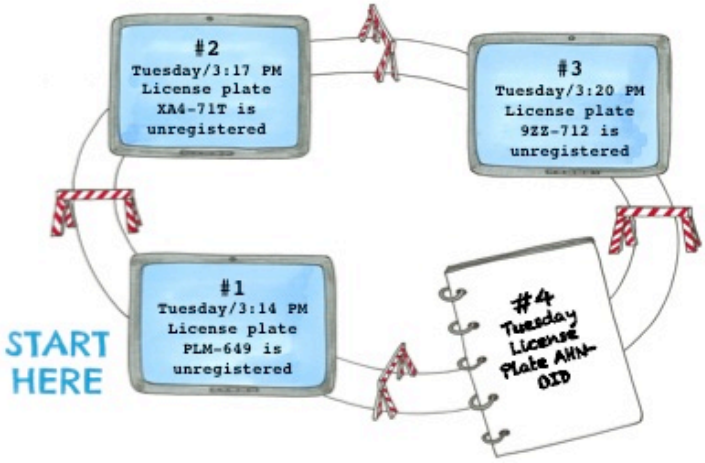
ACT I Curtain - *No attempt at anonymity* is analogous to the subject car having its own license plates on such that the identity of the car, the car's driver and the address of the residence of the driver are known as soon as it drives thru the first traffic light with ALPR capabilities; the time that the car goes through each of the three traffic lights would also be recorded.





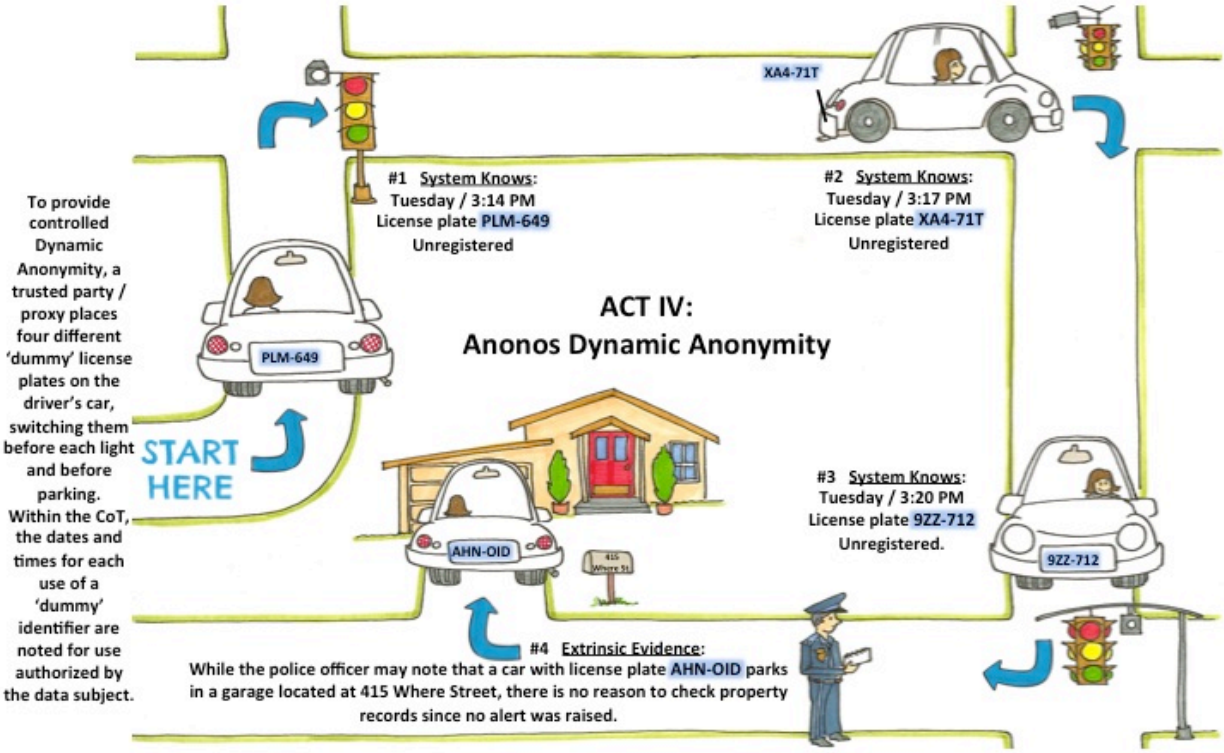
ACT II Curtain – ***Static anonymity*** is analogous to putting 'dummy' license plates on the subject car. While the identity of the car would not be known at the time it drove thru the first (second and third) traffic light it would be known that the same car went through each of the three traffic lights and the time it drove thru each light would also be recorded. At least as early as 2000, experts like Latanya Sweeney, former Chief Technologist at the U.S. Federal Trade Commission, noted in her Carnegie Mellon University paper entitled *Simple Demographics Often Identify People Uniquely* the weakness of static identifiers to provide effective anonymity. More recently, James Turow noted in his 2013 book entitled *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth* that this is particularly the case "when firms intermittently add offline information to the online data and then simply strip the name and address to make it 'anonymous.'."

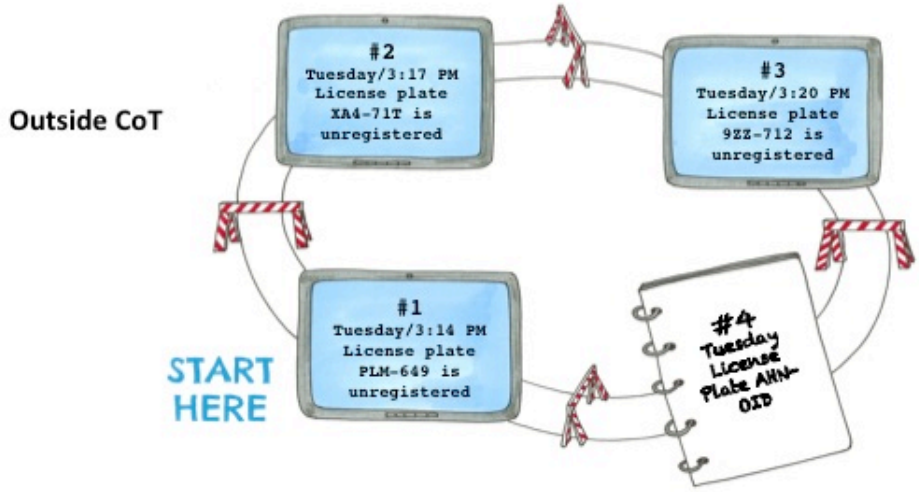




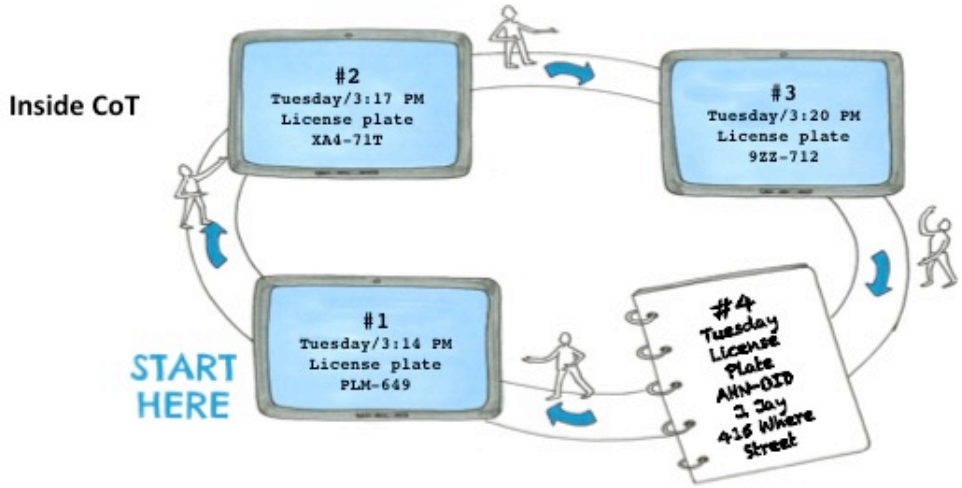
ACT III Curtain – Efforts at "***scrambled anonymity***" like cvdazzle.com, rednoise.org/adnauseam and torproject.org provide a level of anonymity for data subjects, but at the cost of loss of personalized offerings made possible by ongoing advances in technology (e.g., the Internet of Things (IoT), personalized medicine, etc.) and potential loss of social accountability.

For example, a June 10, 2014 post on freedom-to-tinker.com highlighted the rise of a political and ethical theory of obfuscation as "…a strategy for individuals, groups or communities to hide; to protect themselves; to protest or enact civil disobedience, especially in the context of monitoring, aggregated analysis, and profiling…to make life a little bit harder for the powers that be and to protest the surveillance regime...and…to signal to their peers that they are conscientious citizens who will not accept the status quo."

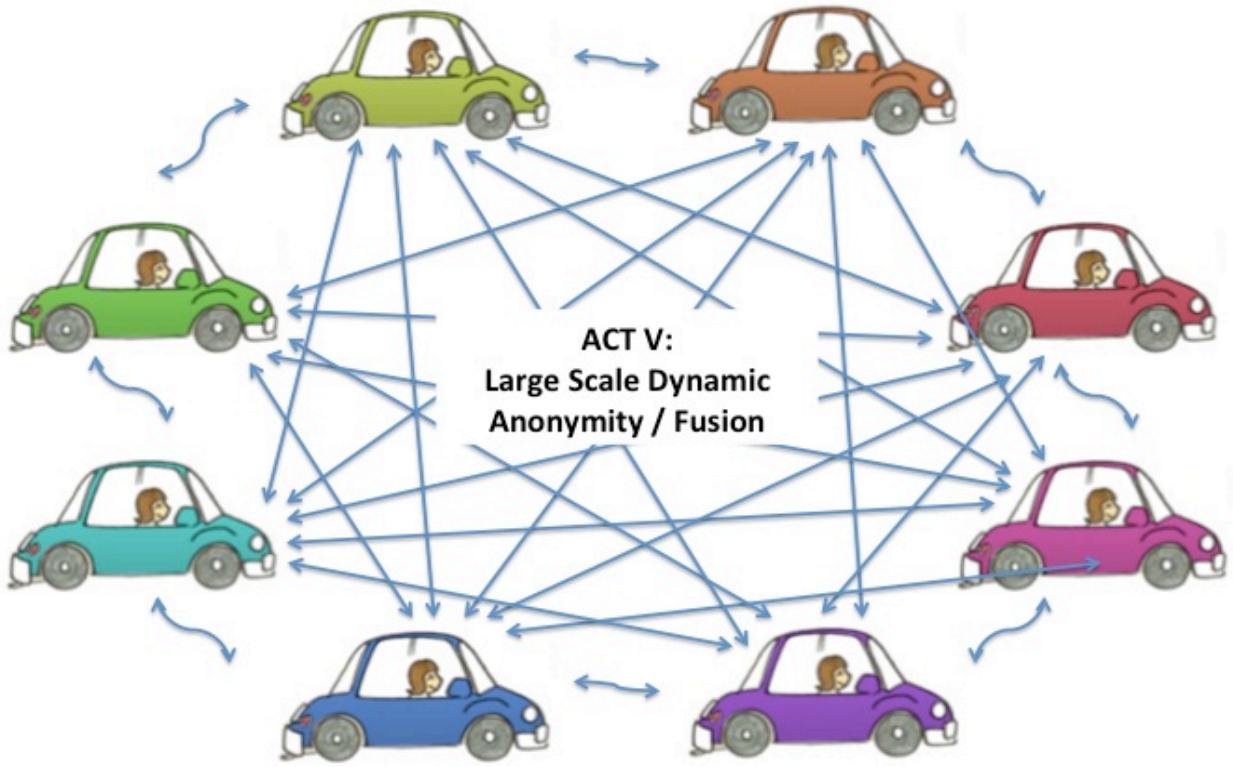




ACT IV Curtain (Part 1) – Outside the CoT with ***Dynamic Anonymity***, the identity of the car is not known at the time it drives through any traffic light; it is not known that the same car went through the three lights; and while the times that a seemingly different car drove thru each light would be recorded, these different times would not be associated with a single car. The use of changing license plates has the practical effect of making it look like the act of driving thru each traffic light is not associated with each other or with a single car. In addition, if someone were to note that the subject car drove into the garage, there is no way to correlate the owner of the private residence to any of the other observed cars / license plates since they would each have had a unique identifying license plate associated with them.



ACT IV Curtain (Part 2) – Inside the CoT with ***Dynamic Anonymity***, controlled data fusion is made possible by providing anonymity for ***Data***, ***Identity*** (of Data Subject / Controller) and ***Context*** (e.g., time, place, purpose) by obfuscating connections between and among the foregoing. The CoT retains knowledge of what dynamically changing identifying license plates were associated with the subject car at each time, place and purpose. This information would be available for uses approved by the user and her trusted party / proxy, which could include access to information by authority figures under established legal practices that include adequate safeguards to protect the rights of the data subject.



ACT V Curtain – combining ***Dynamic Anonymity*** with data ***Fusion*** means scaling Dynamic Anonymity to work with big data, resulting in:

- ***Maximum Return on Investment (ROI)*** for companies / data users; and
- ***Maximum privacy*** for individual data subjects.

Anonos Dynamic Anonymity enables data subjects to benefit from ongoing technological advancements (e.g., the Internet of Things (IoT), genomic research and resulting personalized medicine, etc.) without having to relinquish privacy, security or control. ***This is done by:***

- ***Assigning DDIDs to data / subjects / actions;***
- ***Retaining information regarding association of DDIDs with data / subjects / actions; and***
- ***Providing data subjects and / or trusted parties with deterministic control over who has access to / use of association information***.

Appendix C

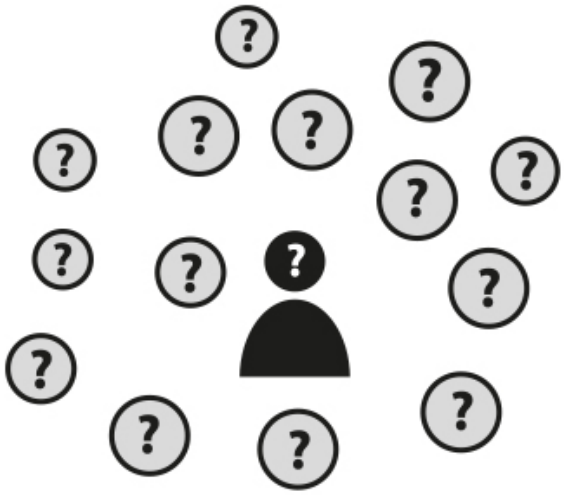
Dynamic Anonymity:
De-Identification without De-Valuation

"De-identification" techniques traditionally used in certain circumstances (e.g., HIPAA or health related circumstances) to protect data privacy are largely defensive in nature – e.g., a series of masking steps is applied to direct identifiers (e.g., name, address) and masking and / or statistically-based manipulations are applied to quasi-identifiers (e.g., age, sex, profession) in order to reduce the likelihood of re-identification by unauthorized third parties. This approach often results in a trade-off between protecting against re-identification and retaining access to usable information.

Traditional De-Identification – Challenges

Traditional approaches to De-identification suffer from:

1. Mistrust of resulting data due to 'noise' and inaccuracies purposely introduced to de-identify data so it cannot be re-associated with data subjects but which can compound errors in subsequent analysis;
2. Increasing risk of re-identification of data subjects due to increasing 3Vs associated with Big Data:
   - Volume
   - Velocity
   - Variety

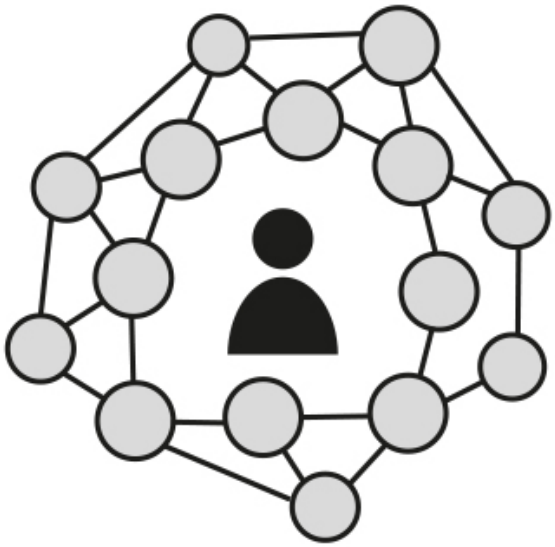


Potential Re-identification of Data Subject Despite De-identification

Accuracy of Data?

Re-Identification – Challenges

1. Increasing 3Vs of Big Data combined with data mining improvements make it easier to re-identify data subjects using quasi-identifiers – i.e., data elements that can uniquely identify a data subject when they are aggregated together and / or combined with extrinsic data.
2. Connections between data elements when aggregated can lead to re-identification of data subjects when using traditional approaches to de-identification.



Re-identified Data Subject

Relationships Between Data Elements

Anonos Dynamic Anonymity has significant offensive value in that the value of information can be retained and leveraged / exploited for authorized purposes, all with a statistically insignificant risk of re-identification of any datum. Dynamic Anonymity rejects the proposition and traditional dichotomy that, in order to minimize risk, one must sacrifice the value of information content. Instead, Dynamic Anonymity minimizes both risk and the amount of information lost, enabling most – if not all – of it to be recovered, but only upon authorization by the Data Subject / Trusted Party, not by unauthorized adversaries / “black hat” hackers.

Anonos Dynamic Anonymity uniquely enables information to be used in different ways by multiple parties in a controlled environment that facilitates unlocking and maximizing the value of data. Dynamic Anonymity maximizes the value of potential business intelligence, research, analysis and other processes while simultaneously significantly improving the quality and performance of data privacy processes.

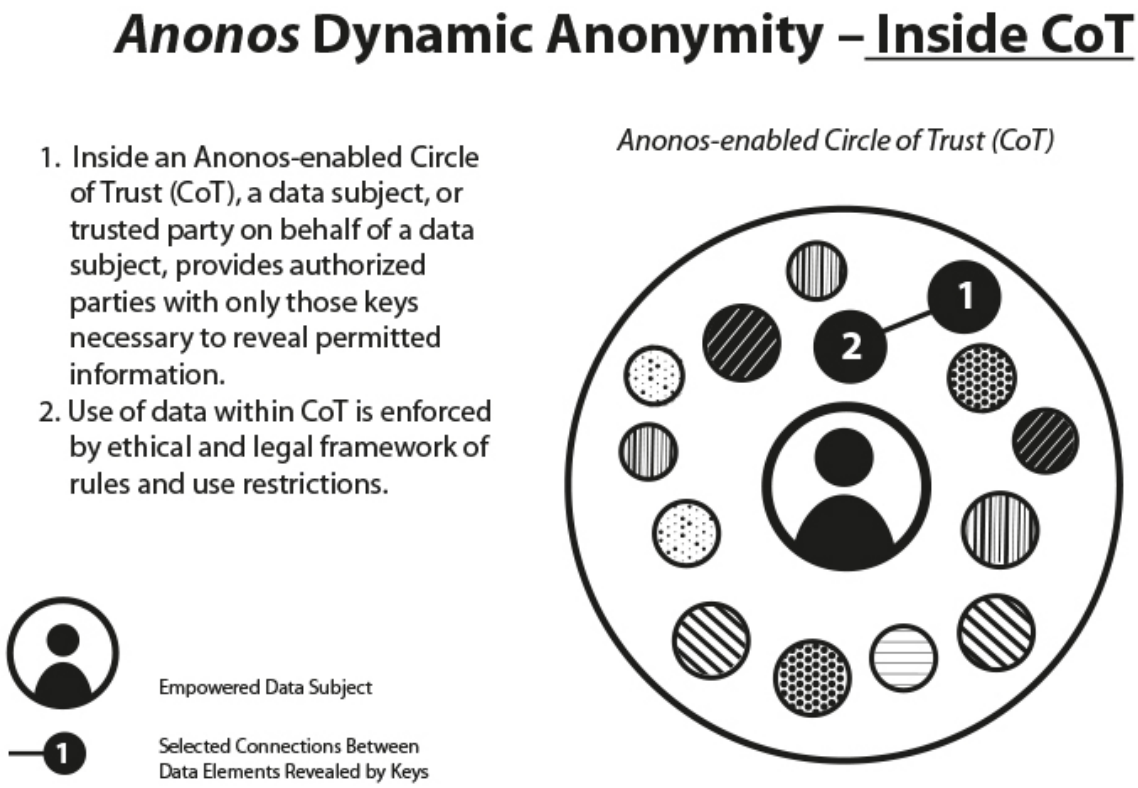
When collected or stored, sensitive data may be “disassociated” from its subject using one or more of the following strategies, none of which incurs any loss in value:

1. Segmentation: Sensitive data may be split into several pieces, by data type, and transmitted and / or stored separately (either in separate Circles of Trust, or using different DDID mapping sets maintained by the same Trusted Party) so that each piece, alone, yields no Personal Data.

2. ID replacement: Static identifiers can be replaced with dynamically changing and re-assignable DDIDs obscuring the relationship between data and the Data Subject to which that data refers.

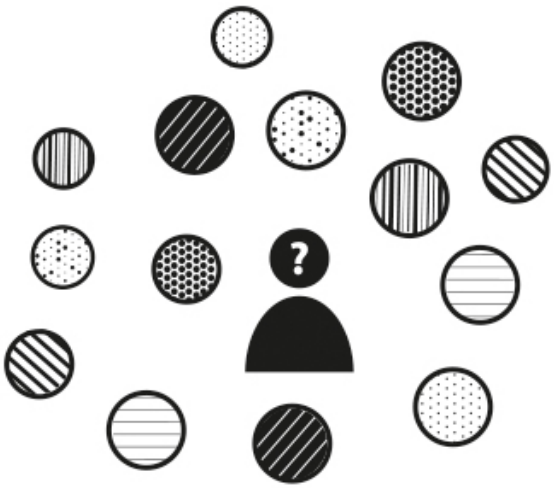
3. Obscuring: data values and data type indicators may also be replaced with DDIDs.

The DDIDs associated with these operations are stored within an Anonos-enabled Circle of Trust (CoT); the original data may thus be reconstituted by reversing these transformations, but only with the cooperation of the CoT itself, and thus only when granted such permissions by, and / or on behalf of, the Data Subject.



Anonos Dynamic Anonymity – Outside CoT

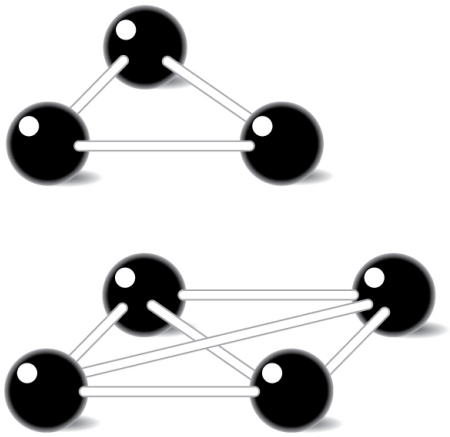
1. Outside an Anonos-enabled Circle of Trust (CoT), Anonos Dynamic Anonymity effects dynamic data obscurity by obscuring connections between and among data elements.
2. Outside the CoT, connections between data elements are not known so they do not re-identify data subjects.



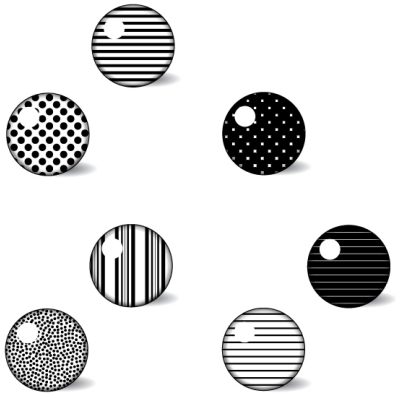
Unknown Data Subject

Unconnected Data Elements

In the first figure below, the different nodes represent data elements related to two different Data Subjects that are capable of being tracked, profiled and / or analyzed by third parties because they can be associated with, and / or re-identified for, each of the Data Subjects. The second figure below presents a simplified visual depiction of the same data elements that can be retained with Anonos Dynamic Anonymity without loss of context necessary to support beneficial “big data” applications; this can be achieved by obfuscating connections between each of the Data Subjects and the data elements in a controlled manner by means of an Anonos-enabled Circle of Trust (CoT).



Non-Obscured Data Elements



Obscured Data Elements

Appendix D

Anonos Co-Founder Bios

Who is behind Anonos?

Anonos aims to interact with industry, privacy and security experts to help address major privacy and security problems encompassing a global scope. This approach reflects the vision of the co-founders of Anonos, 13-year business partners Gary LaFever and Ted Myerson, who believe innovative applications of technology, like Anonos, can facilitate market changes that address the needs of disparate stakeholder groups – including individuals, commercial and not-for-profit organizations, countries and regulators.

Gary LaFever, Co-Founder

Gary is a solutions-oriented futurist with both a computer science and legal background. His combination of technical and legal expertise enables him to approach issues from both perspectives.

- Prior to Anonos, Gary was co-founder at FTEN, a company that revolutionized global financial securities markets by enabling real-time risk management by aggregating together seemingly unassociated data elements to reflect real-time, consolidated financial positions. NASDAQ OMX acquired FTEN following the May 6th "Flash Crash," when the Dow Jones industrial average briefly plunged nearly 1,000 points erasing $1 trillion from the U.S. financial securities markets. This enables NASDAQ OMX to provide technology tools to global exchanges for managing systemic risk in financial securities markets.

- While a NASDAQ OMX executive, Gary co-founded FinQloud, the financial industry big data initiative between Amazon Web Services (AWS) and NASDAQ OMX. FinQloud was the recipient of the Wall Street Letter WSL 2014 Institutional Trading Awards as the Best Cloud Solution.

- Gary is a former partner at the major international law firm of Hogan Lovells, where he specialized in helping emerging technology companies achieve strategic and financial goals in the context of applicable laws, policies, rules and regulations.

- Gary began his professional career at Accenture - the multinational management consulting, technology services, and outsourcing company, following receipt of his undergraduate degree in computer science.

Ted Myerson, Co-Founder

An inventor and visionary with the insight to "see what other people don't see," Ted has a proven record of converting inspiration and innovation into highly profitable businesses.

- Prior to co-founding Anonos, Ted was the founder and CEO of FTEN, a groundbreaking company developing innovative market risk management solutions driving new levels of market integrity. The landmark SEC Market Access Rule 15(c)3-5, sometimes referred to as The FTEN Rule, requiring real-time, cross-market-risk management to improve market integrity was made possible by FTEN technology.

- At FTEN, Ted spearheaded numerous innovations and achievements that led to FTEN's nomination for the 2009 National Medal of Technology and Innovation (NMTI), the United States' highest honor for technological achievement bestowed by the President on America's leading innovators; recognition as an *Inc. 500* 'Top 50' fastest growing company / fastest growing software company two years in a row; and being named a Crain's New York Business "Best Place to Work" in 2010.

- Under Ted's leadership, NASDAQ OMX acquired FTEN in 2010. After the sale, Ted was named Global Head of Access Services at NASDAQ OMX where he managed a division overseeing 16% of total revenue, roughly $250 million in 2012, and 12% of corporate profit.

- Ted was named as a 2010 New York Enterprise Business Report "Game Changer."

Patents Awarded to Gary LaFever / Ted Myerson

2014

- US 8,788,396 - Big Data Cloud Computing System
- US 8,738,479 - Big Data Categorization System

2013

- US 8,489,496 - Risk Management System
- US 8,433,641 - Time Sensitive Big Data Analysis System

2011

- US 8,010,442 - Cross-Market Big Data Management System

2010

- US 7,778,915 - Real-Time Risk Management System

For more information, please contact INFO@ANONOS.com