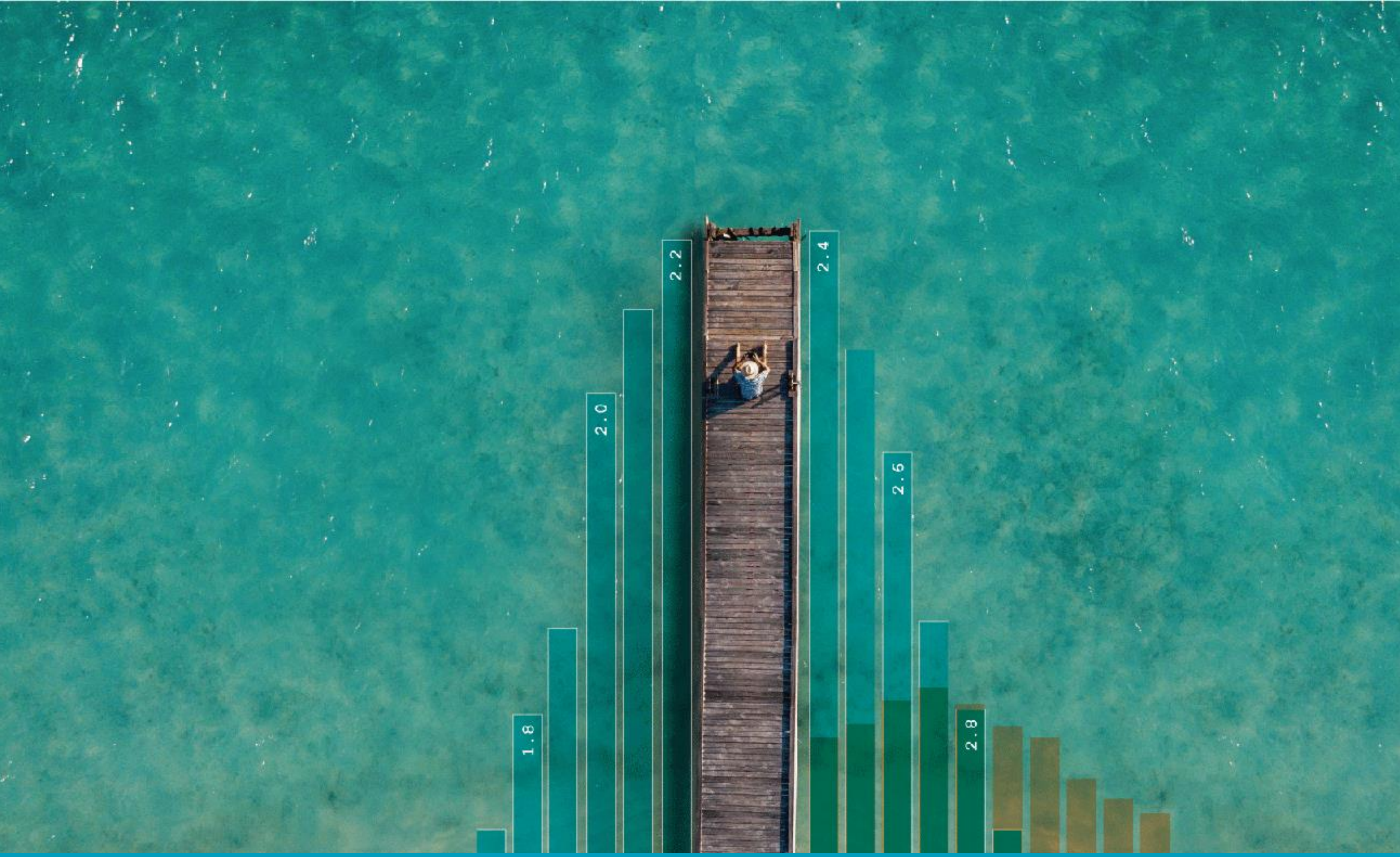
TriNetX

Explore real world, real-time global data



SUMMARY OF THE ASSESSMENT OF TRINETX DATA PRIVACY PRINCIPLES THROUGH AN EMPIRICAL ANALYSIS

Prepared by: Bradley Malin, Ph.D.
brad.malin@gmail.com

December 8, 2020

Abstract
This document summarizes the data privacy principles applied in the TriNetX Platform and their relationship to regulatory requirements.

TriNetX is the global health research network enabling healthcare organizations, biopharma, and contract research organizations (CROs) to collaborate, enhance trial design, accelerate recruitment, and bring new therapies to market faster. The TriNetX Platform is a software application that biopharma, healthcare organizations, and other researchers use to access real-world data to perform exploratory or hypothesis-evaluating analyses to generate real-world evidence. TriNetX supports networks using a combination of hardware appliances maintained by a multi-tenant Software as a Service (SaaS) system. Throughout this document, the combination of the hardware appliance and software system will be referred to as the TriNetX Platform.

The data to be provided via the TriNetX Platform is generated by covered entities, as defined by the Health Insurance Portability and Accountability Act of 1996 (HIPAA). As such, the patient-level information is protected by the regulation. Still, the HIPAA Privacy Rule permits covered entities to disclose, as well as reuse such information, in a secondary fashion [Safran 2007], expressly without the consent of the patients when data are deemed not to be individually identifiable, or "de-identified". It is important to recognize that once data has been de-identified, it is no longer subject to the oversight of the regulation.

The Privacy Rule itself provides an explicit definition of de-identified data, as well as two options for implementation to meet the definition: i) Safe Harbor and ii) Expert Determination. The data to be shared through the TriNetX Platform is longitudinal in nature, and both the relative time between events associated with patient-level encounters, as well as the recency of such events, is critical to ensuring that the semantics of this information are preserved for various application scenarios (e.g., clinical trials recruitment). In this respect, TriNetX aims to retain information about patients that pertain to time periods as short as one day. As described in greater depth below, this window of time is smaller than that which is permitted by the Safe Harbor implementation under the HIPAA Privacy Rule. Thus, the TriNetX Platform intends to make data accessible in a manner that satisfies the Expert Determination implementation model.

TriNetX hired this consultant to review and assist in the refinement of the data protection principles for sharing patient-level data through the TriNetX Platform. In doing so, the consultant reviewed the system design, policies, and workflows to ensure that

1) the risk of re-identification was sufficiently small to meet the requirements of the de- identification requirements of the HIPAA Privacy Rule via the Expert Determination implementation and
2) the exposed data did not violate the principles of site-level protections promised by TriNetX. To assist in this analysis, the consultant reviewed statistics derived from over 100 million patients, contributed by over 80 institutions that may make their data available through the TriNetX Platform.

The purpose of this analysis is to review the capabilities being provided by the TriNetX Platform. This capability allows various parties to download and access de-identified patient records. The patient records are provided by covered entities and made available through the TriNetX Platform. A user with access to the TriNetX Platform can define the patient cohorts of interest from Data Networks that are made available to them. Different organizations and different users have access to different Data Networks depending on their role and access privileges. A Data Network may be the data for a single covered entity or composed into various groupings of covered entities and presented as a single Data Network.

It was noted TriNetX employs a combination of legal-, technical-, economic-, and statistical-based controls to mitigate the risk or re-identification. From an economic perspective, third party users of the data (i.e., not the covered entities studying their own data) supplied through the TriNetX Platform will pay a monetary sum for access. Research has shown that costs can serve as deterrents to potential would-be adversaries and mitigate the chance that re-identification attacks will be perpetrated. From a legal perspective, there are several contractual agreements that may be invoked to mitigate risks. First, all providers of data enter into a Healthcare Organization Network Agreement. Under this agreement, TriNetX commits to protecting and keeping confidential all information it comes across. TriNetX employees who violate the terms of service, could be further subject to civil penalties associated with the violation of corporate computer policies. Second, in the context of a collaborative network, the users of the TriNetX Platform are all providers of data. In this setting, each user of the TriNetX Platform could agree to terms of service, which prohibits the misuse of data in the system, including attempts at re-identification. Third, when the network is opened to include external entities (e.g., a pharmaceutical company), attempts of re-identification are again expressly prohibited.

From a data-based perspective, the data shared through the TriNetX Platform is attenuated to ensure that it does not include sufficient information to 1) facilitate re-identification or 2) allow for the determination of which covered entity contributed which specific information about a patient beyond a certain degree.

The primary ways by which TriNetX protects data is: 1) a minimum threshold on the number of participants when returning aggregate query results and 2) data obfuscation when sharing individual- level records. When thresholding and obfuscation is invoked, data is manipulated at three points: 1) the total count of patients, 2) the site-level counts, and 3) the subtotal of patients who satisfy a specific (clinical) term of interest. The total count of patients is thresholded to ensure that no query directly returns a value smaller than 10, a value that is based in the best practices of a variety of federal and state agencies disseminating data for a variety of purposes, such as public health evaluation. The spirit of these protections was also invoked for generating geographic heatmaps of events.

When data is shared at the individual level, TriNetX employs a collection of obfuscation strategies, which include:

1) limiting geographic regions to no smaller than nine United States Census Divisions,
2) removing or transforming standardized billing codes that are indicative of rare or publicly reportable events (e.g., terroristic activities),
3) limiting dates of death to month and year only,
4) obfuscating dates of birth and events that may lead to the determination of such a date,
5) suppressing additional aspects about an individual's persona, such as the primary language spoken and
6) truncating the tails of distributions associated with physical characteristics to mitigate the direct revelation of outlying individuals.

A statistical analysis of the data was performed using both data from TriNetX and the U.S. Census Bureau to show that the likelihood of a patient re-identification was sufficiently mitigated to satisfy the requirements of the HIPAA Privacy Rule.

The consultant is an individual with appropriate knowledge of, and experience with HIPAA de- identification methodologies as well as generally accepted statistical and scientific principles and methods for rendering information not individually identifiable. The expert attests that the risk is very small that the records received could be used, alone or in combination with other reasonably available information, by the anticipated recipient to identify any individual. The expert has summarized the methods used to make this determination, as well as the results of the analysis in this document, with further details of the assessment available upon request. The expert's evaluation is for the practices associated with the TriNetX Platform and associated data sharing procedures according to their design as of December 8, 2020. In the event that there is a change committed to the design of the TriNetX Platform that influences the grounds upon which this determination is based, the system will require reassessment.

DocuSigned by:
Bradley Malin
3451567A92564D6...

2020-12-08

Bradley Malin, Ph.D.
Consultant

Date

Information About the Expert
Bradley Malin, Ph.D. is the Accenture Professor of Biomedical Informatics, Biostatistics, and Computer Science at Vanderbilt University, where he founded and currently directs the Vanderbilt Health Information Privacy Laboratory. He is an internationally recognized expert on data privacy and has served on national advisory committees for the National Academies regarding the management of data from electronic medical record systems and biorepositories. He is an appointed member of the Technical Anonymisation Group of the European Medicines Agency and the Board of Scientific Counselors of the National Center for Health Statistics (NHSC) of the U.S. Centers for Disease Control and Prevention (CDC). He has consulted on de-identification solutions for numerous industrial, not-for-profit, and government agencies. His research has been published through over two-hundred peer-reviewed articles, portions of which have been cited in the Federal Register, Congressional briefings, and popular media outlets such as Nature News, Scientific American, and MIT Technology Review. Among various honors, he received the prestigious Presidential Early Career Award for Scientists and Engineers (PECASE) is an elected fellow of the National Academy of Medicine (NAM), the American College of Medical Informatics (ACMI), the American Institute for Medical and Biological Engineering (AIMBE), and the International Academy of Health Sciences Informatics (IAHSI). He received a bachelor's in biological sciences, master's in public policy and management, and doctorate in computer science, all from Carnegie Mellon University.

Additional information can be found at http://www.hiplab.org/people/malin.