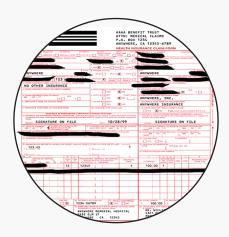# DON'T REDACT!™

ANONYMOUS DATA DOESN'T HAVE TO BE DIFFICULT OR EXPENSIVE ANYMORE

SENSITIVE DATA USE?

Don't Redact!™ is a multi-source anonymization engine. The engine consumes data sources with sensitive information and outputs data safe for use in testing or analytics. As opposed to simply redacting, masking or suppressing the sensitive data, Don't Redact!™ maintains personhood and record integrity in the data that has been output. While names and other identifiers have been changed in the output data, they are always changed in the same way in a given customer instance. This not only allows data coming from the same source over time to be matched with previous records, but allows data coming from a completely different source to be matched with records which have previously been anonymized.

TODAY'S COMMON THREATS

A decade ago, the need for anonymized data was infrequent and irregular enough that when required, a custom designed and developed solution was hard coded specific to the situation being faced. This difficult and expensive path meant many projects either proceeded with real data or were abandoned before they started. Today, this need is far more common and the threat of breach – even in mundane scenarios – is very real.

Tools available for multi-source anonymization should be getting easier to use, cheaper, more capable and simply more available. Instead, the solution is increasingly larger teams dedicated to more complex, expensive, hard coded solutions – even when they are "product based". Hard coded solutions that are difficult to re-use next month or next quarter when a new environment needs to be tested – let alone when an additional data source is added.

Key fields like name can be anonymized in the same way even when they come from completely separate sources.

ANONYMIZATION ENGINE

Faster
Create new data sets & sources

Easier
Master the tool in hours

Powerful
Multi-source data anonymization

Cheaper
Starts at $25,000

DON'T REDACT!™

While companies turn to Don't Redact!™ for a variety of uses, the most common are: Generating Synthetic Data for testing and Complex Analytics studies which require ongoing record integrity.

TEST DATA SETS

When bringing up a new line of business system or reporting environment, many resources are involved. These might be Internal IT resources, Data Scientists, external vendors, contractors or a host of others. Giving access to sensitive Personally Identifiable Information (PII) increases the potential the data will be exposed, breached or used for purposes other than intended. However, using masked, randomized or redacted information as a test set does not produce real world scenarios. Testing might either be impossible because the data is too far off (e.g. all John Doe's make the data massively duplicated) or not uncover problems that real data will expose – after go live.

Having data that looks real in every way is important. If a new transaction management system is being tested, transactions over time made by the same person shouldn't be randomized. The data should always be anonymized the same way so that those transactions can be matched to the same person and account. In addition, if the transaction management system being tested depended on data from a master customer system, that data must not only be anonymized as well, but also be done in the exact same way as the transaction management system – otherwise no transactions will ever match a customer.

COMPLEX ANALYTICS

The tools available for gaining complex analytics insights are not only powerful, but capable of producing amazing insights which lead to real world improvements in systems and outcomes. These tools however are only as good as the information available to them. Gaining insights into complex issues such as cancer incidence or long term clinical outcomes takes data from many sources collected over time. Masking, redacting or suppressing PHI (Protected Health Information) place a hard limit on the types of trends and insights even the most advanced analytics environment can find. Studies that can take place often have such limited data sets that hoped-for results are not realized.

Today, a false choice is often presented to those looking to find real, actionable, original and valuable insights – either use full PHI data in the study and hope to protect yourself through data use agreements or use data that has been so obscured by masking data the goals of the study will never be realized. Often, the result is project paralysis.

GET IN TOUCH

For more information, to purchase or to find out what else we do:

sales@theEDIproject.com
760-602-4394
www.theediproject.com/anonymization.html

A COMPLETE ANSWER

Don't Redact is a powerful tool for anonymization of sensitive data from multiple sources. It depends on the user to know their data, identify which fields need to be preserved, randomized, anonymized or suppressed. It is simple to learn, can be quickly deployed and delivers consistent results for the life of the project. The system starts at $25,000 the first year and is $2,500 per year after that.