

Privacy Analytics Lexicon

Enabling Unstructured Data for Secondary Use

80% of healthcare data is unstructured. This data is critical to deriving new insights, innovation and knowledge for research hospitals, insurance and medical claims providers - among other industries. Accessing and analyzing this data, while ensuring patient privacy is properly protected, will improve healthcare outcomes and drive operational efficiencies.

To meet the growing demand, **Privacy Analytics Lexicon** ensures a HIPAA-compliant approach to redact and de-identify unstructured data. This software enables medical researchers and data analysts to access valuable unstructured data, while allowing data managers to safeguard personal information when it is used for secondary purposes.

HIPAA Compliance for Unstructured Data

The growing volumes of personal information in digital formats heighten organizations' susceptibility to data breaches and regulatory scrutiny. With **Privacy Analytics Lexicon**, privacy and compliance officers strengthen the protection of their organizations' data assets by extending de-identification to unstructured data. With a standardized, enterprise-wide approach, privacy and compliance officers can:

- Ensure company-wide de-identification practices that are compliant with HIPAA and other legal requirements, as well as internal organizational policies and procedures governing privacy and personal information;
- Extend the practice of de-identification to unstructured formats residing in electronic health records, medical devices, clinical

notes, discharge summaries and other data sources, to optimize protection and coverage of data for secondary purposes; and,

- Increase the value of underlying data assets with the ability to capitalize on de-identified unstructured information for analysis.

Gain Analytic Utility and Insight

Privacy Analytics Lexicon redacts and de-identifies personal information, such as names and dates of birth, found in physician notations in structured databases, medical devices' text fields, or XML. It can de-identify MS Office files, text fields in unstructured databases, PDF files and more.

Privacy Analytics Lexicon allows data custodians to configure the redaction and de-identification of unstructured data, preparing it for secondary use and analysis, by:

- Discovering and annotating personal information residing in multiple text formats and fields in RDBMS sources, including ID's, credit cards, driver licenses and medical codes;
- Improving the quality of de-identified data;
- Allowing for better insight into temporal and geospatial data by preserving the granularity of dates and zip codes; and,
- Evaluating the measurement and tuning of precision and recall by comparing pre-determined samples of a dataset.

Integrated Solution

Privacy Analytics Lexicon provides users with a highly scalable solution that, through its API, automatically integrates with customers' IT environments. It can extract, de-identify and index



personal information in text fields residing in standard database tables. Some technical benefits include:

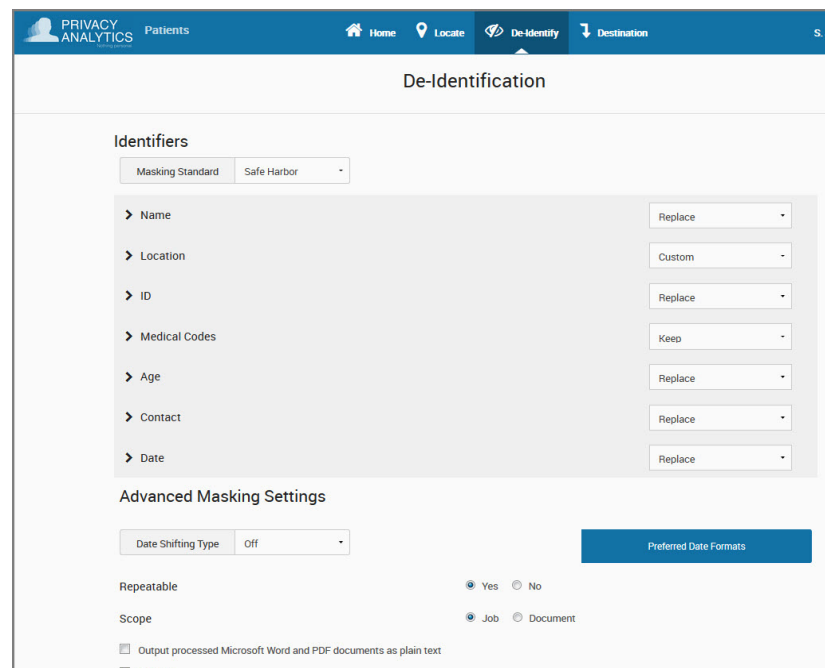
- Custom annotation types and discovery options for increased accuracy
- Flexible de-identification options for repeatable masking and date shifting
- Highly scalable web-based architecture with user management and REST API's

Used in conjunction with **Privacy Analytics Eclipse**, they allow analysis of both structured and unstructured data for the most complete 360 degree view of the patient.

For medical researchers and data analysts, **Privacy Analytics Lexicon** enables them to:

- Automate the de-identification of unstructured data from multiple sources to gain richer analytic value and insight;
- Mitigate the risk of re-identification by detecting personal information, while also applying different de-identification techniques to preserve its relative analytic quality; and,
- Enhance the value of data assets by maintaining the relationships within de-identified data for more granular, higher quality analyses.

Privacy Analytics Lexicon: the proven, responsible way to share unstructured health data.



Privacy Analytics Lexicon applies different HIPAA-compliant de-identification techniques, moving beyond simple redaction. Users can quickly and easily create a compliant dataset while preserving the analytic value of the data.