

## CASE STUDY



# Delivering high-quality, compliant data for analytics



Member-owned, not-for-profit indemnity organization supporting more than 300,000 doctors, dentists and healthcare professionals around the world

### Industry

Healthcare, Legal

### Solution

Data Protection and Analytics

### Key benefits

- Quality analytics
- Decreased costs
- Privacy compliance

## Opportunity

Medical professionals rely on professional organizations to help them understand and navigate the ongoing challenges of regulation and risk management. One professional organization wanted to assist its members by delivering deep analysis of key knowledge found in 100 years' worth of information related to liability claims.

The case material – medical records, associated documentation, and legal findings – is a rich and valuable resource for data mining. However, it contains a lot of very sensitive personally identifiable information (PII) and cannot be stored long term due to privacy regulations such as GDPR. And, the material itself is challenging to analyze, since the documents are mainly unstructured – emails, letters, text documents, etc.

The organization wanted to create a redacted content store allowing them to search and analyze all material without personally identifiable details. This content store needed to incorporate semantic meaning, to make it easier and faster for analysts discover new insights across decades of data.

Connecting facts and their meaning would allow the organization to answer important questions like “what claims are we likely to encounter and how do we educate our members away from risk?” and “when a claim is made, what is our historic win rate? What factors influence this? Should we pay or contest the claim?”

**Large-scale analysis of redacted documents allows the organization to identify trends in their indemnity cases, and take action to mitigate potential issues.**

## Results

The organization chose the MarkLogic data platform with Semaphore Semantic AI to redact sensitive data and structure documents to enable predictive analytics.

The solution has enabled the organization to secure personally identifiable information to comply with regulations, deploy a rich and robust classification strategy based upon SNOMED, and leverage key knowledge for analytics from information that was once unavailable.

- Semaphore's Knowledge Model Management module enriches and manages a model that incorporates the SNOMED ontology. SNOMED, a comprehensive, standardized, multilingual vocabulary of clinical terminology, helps in the understanding of medical content by reducing the variability in the way data is expressed and harmonizing it in a form that can be exploited for analytics.
- The enriched model is published and used by Semaphore Classification and Language Services to automatically examine each information asset and appropriately identify key items such as name, title, address, medical number, birth dates, and other identifying facts, as well as apply classifications from SNOMED and ICD10. This latter enrichment process replaced the organization's manual ICD10 classification processes with automated, next-generation SNOMED classification which provides much more granular insight.
- The Semaphore Fact Extraction Framework service created a document fingerprint that combines Natural Language Processing (NLP) entity recognition with identification of contextual facts to identify medical PII in the organization's diverse set of content types. Sensitive content is identified, marked for redaction, and managed according to current regulations.

The Semaphore model and metadata are used to classify and tag medical information found in the textual documents that was once unavailable to the organization. Models, data, and metadata – facts and their meaning – are securely stored and managed within the multi-model MarkLogic data platform, where they can be searched and analyzed using both semantic and free text queries.

A two-week proof of concept project proved the platform's ability to redact, classify, and search content. Now, large-scale analysis of all case-related documents allows the organization to identify trends in their indemnity cases and take action to mitigate potential issues. The organization is decreasing costs while remaining compliant with privacy regulations.