

Client Case Study: Better safety outcomes for the construction industry



Desensitizing 800,000 documents to enable safety learning sharing between construction partners. The Data X-Ray helped them accomplish this in **1 day instead of the several years** it would take to do the same manually.

Challenge

One of the top engineering firms in the world, Wood Group, was tasked with identifying safety incident typologies with construction partners. The goal was to be able to compare incidents across multiple organizations to understand how and why accidents happen. Unfortunately most of these safety records had sensitive and/or personal data in them, which made sharing of that data problematic. Across almost one million records, desensitizing these reports manually would have taken years.

Solution

Wood engaged Ohalo as an expert in the field of automated unstructured data redaction. The Wood solution needed to i) find where sensitive data is within large unstructured data sets, ii) annotate that data

with sensitive data tags according to a custom redaction policy, and iii) automatically remove that sensitive data from the datasets.

Wood selected Ohalo's Data X-Ray due to the previous work Ohalo had been doing with the UK's health and safety regulator, HSE.

The Data X-Ray had all of the requirements that were necessary to achieve the business case on the first one-off basis and also give Wood the capability to make their health and safety data more useful over time since the Data X-Ray has built-in features not only for redacting data at scale but also for data search and classification.



Process at 100,000s of words per second

Different from a human, the Data X-Ray can read 100,000s of words per second, unlocking use cases not possible with manual processes

Achieve near-human anonymization accuracy in unstructured data

Given enough training, the Data X-Ray can accurately redact up to 99% of records processed.

What was done?

Wood wanted to get going immediately so they opted for the Data X-Ray no-integration cloud option using the Data X-Ray's in-built drag-and-drop storage solution.

Since Wood opted for the no-integration cloud option, the initial set up took only a couple of hours. Ohalo staff assisted Wood in initial training and onboarding consisting of:

- Initial data upload to Data X-Ray managed storage ("Dropbox" like drag and drop)
- Building operating model workflows within Data X-Ray for specific redaction requirements
- Configuration of algorithms for redaction
- Casefile creation and management

The initial project was led by Wood's Digital & Technology team with Jason Payne managing and Billy Archbold taking on the operational role of actually ingesting and desensitizing the records.

Wood and Ohalo were able to build a very close working relationship so during the initial phase if any issues arose, we were able to fix them in very short order to keep Wood on schedule with their initial roll out work.

Results

The Wood Group relied on previous work that the HSE has done with Ohalo as an assurance that their documents were being effectively redacted since the data being used was the same type of data and the rules being implemented were also the same.

With the HSE, the anonymization effectiveness of the Data X-Ray was evaluated against manual anonymization. This evaluation was based on 1,998 RIDDOR reports used for the Construction Division RIDDOR dashboard, which was manually anonymised in 2017 and made public.

The standard of assessment used for significant breach is that of the UK's privacy regulator, the Information Commissioner's Office (ICO), and involved making a determination as to whether there will likely be a risk to people's rights and freedoms.

Of the 743 manually anonymised records containing personal data, **213 (29%) were identically auto-anonymised by Data X-Ray and the overall achievement indicated that 99%+ of records had no sensitive information remaining in them.**

The end result for Wood means that they now have a capability to redact records on an ongoing basis to make the safety records more useful while also providing capabilities to search and find desensitized data relevant to onsite activities going forward. At the end of the day this is contributing to getting workers home safely to their families each and every day.

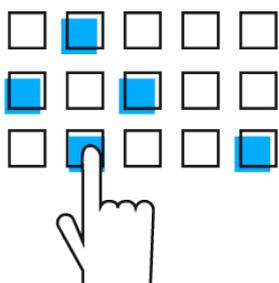
Ohalo's Data X-Ray software allows participating contractors to easily and confidently redact any personal information from their safety records

- Billy Archbold, Wood Group

Data X-Ray for Anonymisation

ENGINEERING YOUR PRIVATE DATA

Automated data ingestion and redaction at scale with the customization required for your data

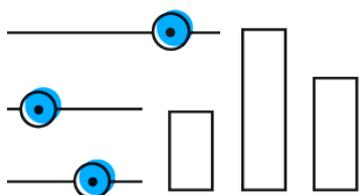
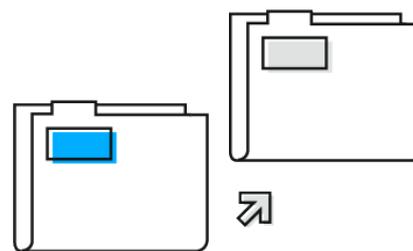


Integrate with native datasources or use Data X-Ray's no-integration cloud options

Companies can operate literally thousands of databases. Data X-Ray integrates to most datasource types from SQL databases to Windows file servers to cloud storage in seconds. Data X-Ray also provides no-integration options using secure managed cloud storage.

Manage separate redaction projects with casefiles and custom classifiers

The Data X-Ray allows you to build out casefiles with subsets of data and overlay custom classifiers onto those subsets of data for ultimate accuracy. This is all available out-of-the-box on day one but can integrate to your existing data pipelines seamlessly through the REST API.



Build your own custom classes to find data that is sensitive in your context

The Data X-Ray rules customization allows a user to define new AI classes, build regular expressions, and match keywords that are sensitive within a particular data pipeline context.