

DATA XRAY

REDACTION AND DESENSITIZATION

Remove personal and sensitive information from your data at scale



Privacy-preserving data science, analytics, and business processes

With scaled datasets, manual desensitization of documents is not an option. Whereas humans can read and comprehend around 10 words per second (if they are fast), the Data X-Ray moves you to a world where you can desensitize data pipelines at scale and at 100,000s of words per second at the click of a button. This gives you the power to scale up and down redaction teams 10,000X faster than you would be able to do in traditional manual processes.

The Data X-Ray enables ongoing privacy-preserving data science by connecting into your data through native connectors or APIs and allowing you to create, deploy, and iterate on custom desensitization algorithms that achieve ever-higher accuracy.

Built into your data pipelines

The Data X-Ray integration is simple with built-in native connectors for data at rest and a full API suite for data in motion and integration to custom applications.

Good-as-human accuracy

The Data X-Ray allows you to

Case study: automatic anonymization and desensitization of health and safety data



The UK government's health and safety regulator, the HSE, needed to anonymize personal and sensitive data from **600,000 free text records** for sharing with third parties to share how accidents are happening around the UK. The task was estimated to take **12.5 person years** if done manually.

The Data X-Ray reduced that to **1 machine day at a 99% accuracy rate**. The result? **4,500x less time at 49x less cost**. This gave the HSE's legal team comfort that sharing their data would not violate the personal data rights of the data subjects in the datasets and ultimately allowed the HSE to share their data with third parties to reduce accidents and fatalities in the UK.

See more in the detailed [case study](#).



mix the latest in natural language processing algorithms with your custom dictionary and regular expression rules. It is possible to achieve manual-level accuracy rates after 3-4 weeks of algorithm iteration. With a UK government client that required highly secure deployment on their environment the Data X-Ray was able to redact sensitive data out of 99.4% of records in *1 day instead of the 12.5 years* that it was estimated to do the same work manually.

Ingest

The Data X-Ray natively reads hundreds of file types as well as database columns. You can ingest any normal type of file or database column as your raw dataset.

Redaction performance reporting and record keeping

Optional redaction reports include statistical reports of what was redacted and why so that legal or privacy teams can review the results and develop a governed process around your redaction or desensitization activities.

Multiple redaction options

In addition to our built in natural language processing algorithms, regular expressions, and dictionaries, you can also use the Data X-Ray's redaction filters if you have a requirement to achieve very high recall on certain sensitive data types.

Export options

In addition to the redacted versions of the documents, you can export raw files as well as unredacted text versions of the data for further downstream processing. You can also choose how to redact (hash, obfuscation, replacement, etc.)

SCHEDULE A DEMO TODAY

www.ohalo.co/demo