# Lloyds Register Safety Accelerator Final Report

## Automatic anonymising and desensitising of health and safety data

Version 5
6 March 2020

Presented by:
Kyle DuPont, CEO
Ohalo Limited

| | |
|---|---|
| **Client name** | Lloyds Register Foundation |
| **Ohalo contact** | Kyle DuPont, kyle@ohalo.co or admin@ohalo.co |

# Introduction

## Overview of challenge

Lloyds Register Foundation (LR) were looking for innovative proposals for solutions that are able to accurately and expediently desensitise and anonymise health and safety information sources. This might include information held in both structured and unstructured data formats, contained in spreadsheets, databases and reports. The latter might include reports in Word documents, pdf's, or any other machine readable file formats, including both short (e.g. a few pages) as well as longer (>50 pages) volumes.

## Proposed solution

The solution calls for a tool that is able to i) find where sensitive data is within large unstructured (and possibly structured) datasets and across various file formats, ii) flag where that particular data might be, and iii) automatically isolate sensitive data from the original dataset.

The Data X-Ray is Ohalo's proprietary automated data discovery and mapping tool which uses machine learning technology for data protection compliance and data governance purposes ("Data X-Ray").

The Data X-Ray features connectors built for both unstructured and structured data sources in all principal file formats to i) extract data from its original format, ii) analyse that data to discover potentially sensitive items, iii) separate sensitive data from non-sensitive data, and iv) output that data to a usable format.

The solution is a service installed within an HSE environment and provided to HSE and its partners to share and redact data with the goals of:

- Sharing data with research partners (in particular, the University of Manchester) in a GDPR compliant way
- Having third party industry partners and other subject matter experts able to share data with HSE in a GDPR compliant way
- Evaluating automatic basic document organization and categorization tasks with machine learning to prepare large data sets for more effective analysis in the future.

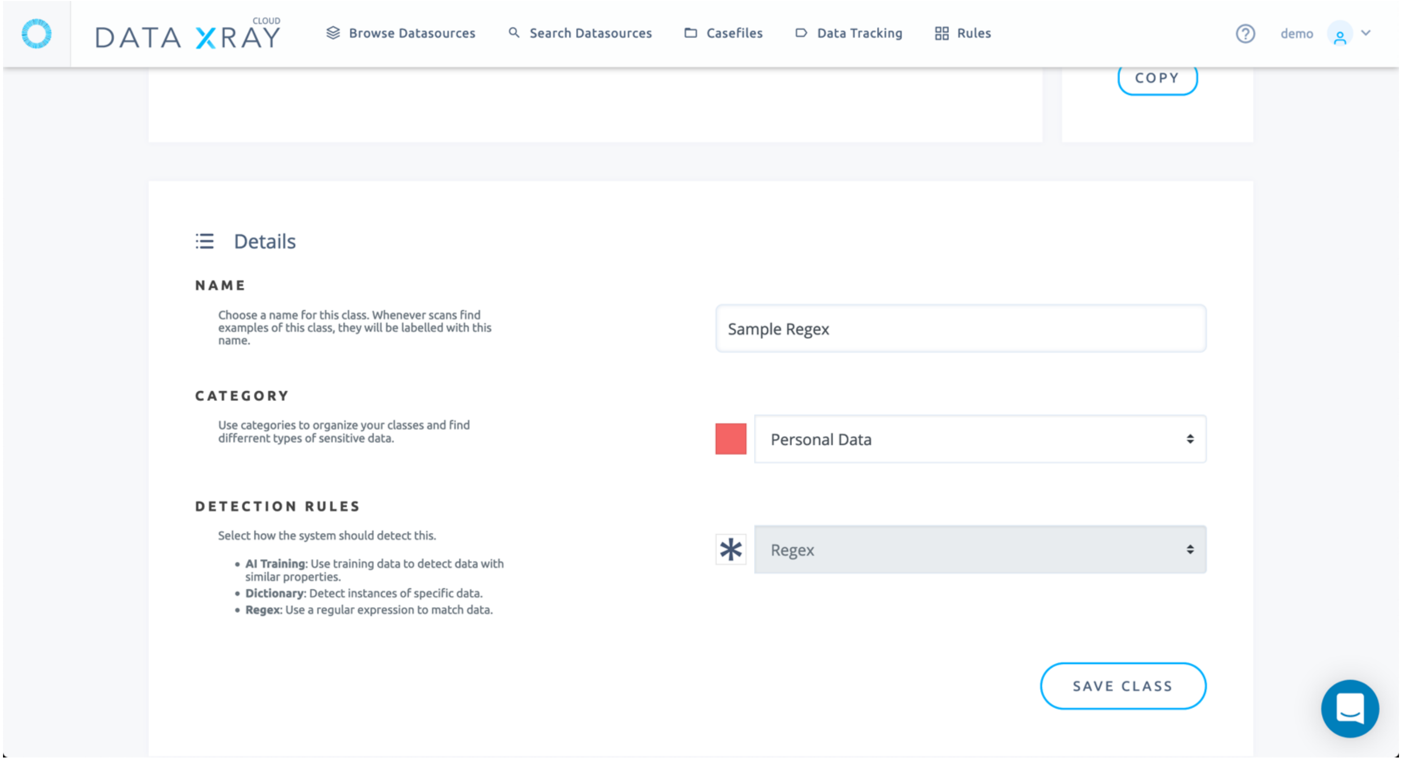
# The Pilot

## What was done?

Ohalo set up a server at HSE's premises with an instance of the Data X-Ray. The server took in data from an HSE data source (HSE RIDDOR report data, details in Results section, below), analysed that data for personal data, and redacted it.

A team led by the HSE data science team looked for any sensitive information that had not been redacted and the ability to personally identify individuals or entities by 'joining up the dots', for instance by linking PII data to public data to infer identity.

Where false or true positives were discovered by the HSE team, the Data X-Ray enabled the team to update the models with example training information to facilitate improved redaction results. Multiple techniques were used to update the models, such as adding new ML classes, dictionaries, and regular expressions and some backend data engineering techniques that involved ensuring word tokens are correctly combined to achieve better token concatenation. As a visual description, the images below explain how an additional regex can be added to the model.



Details

NAME

Choose a name for this class. Whenever scans find examples of this class, they will be labelled with this name.

Sample Regex

CATEGORY

Use categories to organize your classes and find differrent types of sensitive data.

🟥 Personal Data

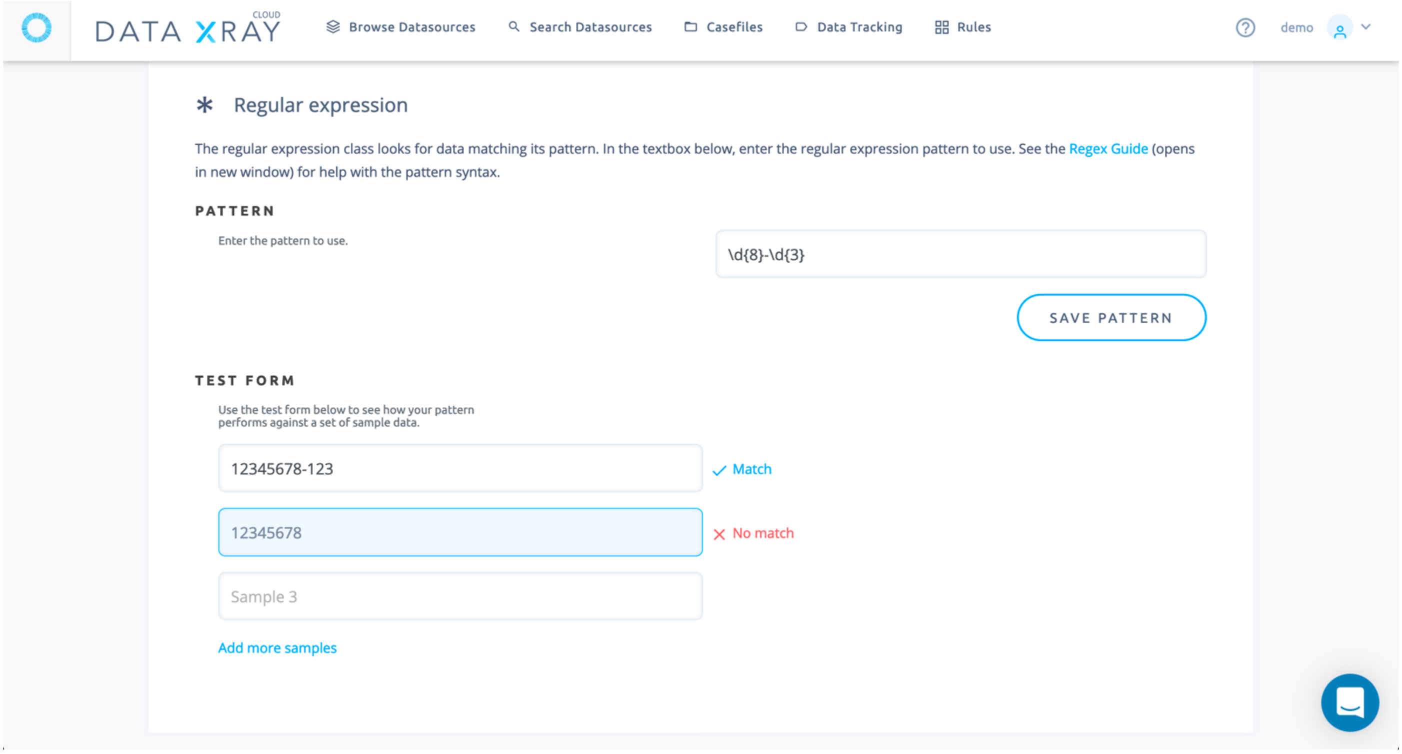
DETECTION RULES

Select how the system should detect this.

- **AI Training**: Use training data to detect data with similar properties.
- **Dictionary**: Detect instances of specific data.
- **Regex**: Use a regular expression to match data.

* Regex

*Image 1: Creating a regex (1)*



* Regular expression

The regular expression class looks for data matching its pattern. In the textbox below, enter the regular expression pattern to use. See the Regex Guide (opens in new window) for help with the pattern syntax.

PATTERN

Enter the pattern to use.

```
\d{8}-\d{3}
```

TEST FORM

Use the test form below to see how your pattern performs against a set of sample data.

| | |
|---|---|
| 12345678-123 | ✓ Match |
| 12345678 | ✕ No match |
| Sample 3 | |

*Image 2: Creating a regex (2)*

Issues identified were usually tractable and resulted in a very high number of records to be appropriately anonymised as to result in a very small percentage of records that would have resulted in a data breach (as below, there are around 10 records out of 1998 that Ohalo considers to be difficult to redact in an automated way with Ohalo's current technology). However, it was identifying these issues in the first place that proved to be a challenge at scale.

To illustrate this point, one of the biggest problems that the team ran into during the engagement was automatically comparing the results of the automated anonymisation to the results of the manual anonymisation to identify issues that need to be fixed and ensuring that the fixes implemented did not inadvertently create other issues. This was first done with manual review of the results but it was soon found to be difficult to do even with the 1998 records that we were comparing against. Therefore the HSE team developed a Python script that automatically compared the results of the Data X-Ray through its API and against the manually redacted records.

Subsequently, an independent team at the HSE graded the accuracy of the redaction and in tandem with Ohalo iteratively improved the accuracy of the models used to redact the data. This involved an iterative process between one staff at HSE and one at Ohalo to identify false negatives and false positives and understanding why those false negatives and false positives occurred. In addition to checking the results with the automated script that HSE developed, another staff at HSE checked and graded the final results.

### Was it on time and on budget?

The project experienced delays at the beginning due to HSE's internal data privacy requirements around data transfer to Ohalo-accessible servers. Therefore the project ran several months over schedule due to those approval processes.

The project was delivered on budget.

## Results

### Did the new solution improve on existing solutions used by HSE?

The effectiveness of Ohalo Data X-Ray anonymisation was evaluated in comparison to manual anonymisation. This evaluation was based on RIDDOR data used for the Construction Division RIDDOR dashboard (https://www.hse.gov.uk/construction-dashboard/) which was manually anonymised in 2017 and made public.

The standard of assessment used for significant breach is that of the ICO was that there will likely be a risk to people's rights and freedoms.[1]

From the 1998 RIDDOR reports analysed 743 contained sensitive. Anonymisation using Ohalo Data X-Ray resulted in 94 retaining some PII of which 19 would be considered sufficient for a significant breach. Assuming that all RIDDORs containing PII would be considered sufficient for a significant breach this has reduced the number of sensitive records by 97%.

Whilst the data cannot be considered fully effectively anonymised, the data sensitivity, and hence impact of a breach and resultant risk, has been reduced very significantly. This and in concert with Data Processing Agreements allowed a better evaluation of the remaining residual risk in data sharing.

Table 1 summarises the final results from the HSE final report.

[1] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/personal-data-breaches/

*Table 1: Record statistics extracted from the Evaluation Spreadsheet*

| Description | No. of records | Comment |
|---|---|---|
| Total No. of HSE RIDDOR records evaluated | 1998 | |
| No. of records manually anonymised (containing sensitive text) | 743 | Note that a small amount of sensitive information was not redacted manually e.g. Brand names |
| No. of records having no sensitive text in need of removal | 1255 | As assessed by manual anonymisers |
| No. of records auto-anonymised by Data X-Ray | 1070 | This figure exceeds the 743 manually anonymised, principally due to over redaction of time period entities alongside dates. |

Of the 1998 records analysed, only 743 needed sensitive text removed (i.e. 1255 records contained no sensitive information).

Of the 743 manually anonymised records, 213 (29%) were identically auto-anonymised by Data X-Ray.

The majority of the differences are due to both under and over redaction by Data X-Ray. However, some minor differences are also due to manual alterations (e.g. spelling corrections) made to the original records during the anonymisation process; and also some differences due to manual under redaction; 69 of the manually redacted records retained elements of 'sensitive' text e.g. tool brand names, first name, company acronyms, motorway name. **This serves to show that even manual redaction is not 100% consistent; in this case individual assessments of sensitivity led to differing decisions with sensitive elements that were not PII.**

There was some over-redaction of text, primarily of time periods. Of greater concern was under-redacted sensitive text, i.e. that which Data X-Ray allowed to 'slip through the net'. Table 2 lists the categories and the number of records retaining sensitive text. (Note that a single record may contain text from multiple categories i.e. there is some overlap within the reported figures).

*Table 2: Under Redaction Details*

| **Category of under-redacted sensitive text** | | **No. of records** | **Percentage (relative to 743 manual anonymised records)** | **Percentage (relative to 1998 total records analysed)** |
|---|---|---|---|---|
| No of Data X-ray **under** redacted records<br>Of which: | | 267 | 36% | 13% |
| | PII<br>(Significant Breach) | 94<br>(19) | 13%<br>(3%) | 5%<br>(1%) |
| | Gender (Title) | 48 | 6% | 2% |
| | Company Name | 42 | 6% | 2% |
| | Location | 83 | 11% | 4% |
| | Date | 34 | 5% | 2% |
| | Reference No. | 31 | 4% | 2% |

Of the 1998 total records, 94 records retained PII, and only 19 were considered to be a significant breach of GDPR, as defined by the method of assessment outlined above. **This serves to show that the Data X-Ray's automated anonymization removes PII resulting in 99%+ of the records being anonymised.**

## Business case and commercial feasibility

The pilot proved that we can successfully redact personal and sensitive health and safety data from unstructured data sets to a very high degree of accuracy. Using this data, HSE will be able to share data with third party researchers, for the furthering of HSE's mission to prevent death, injury and ill health to those at work and those affected by work activities. However it is ultimately up to the discussions at HSE in terms of what an acceptable level of risk would be, given the existing controls in place around their data sharing partners. On a wider basis, this also provides a sound basis for the use of automatic anonymisation technology to share health and safety data between organisations for research as part of the data minimisation process.

The problem set that HSE has: the anonymisation / redaction of 0.6m RIDDOR reports (and over 1m other different types of documents) is not a tractable problem with manual techniques. It was calculated that it would take HSE up to 12.5 person years of time just to start with the 0.6m existing RIDDOR reports, much less the stream of reports that come in every month (up to 10,000/month) and the other documents needing anonymisation.

Additionally, manual redaction was shown to be imperfect, leaving 69 out of 2,000 records with sensitive data remaining in them.

The techniques available with the Data X-Ray allow HSE to anonymise data at a speed and accuracy and cost not possible using manual processes. In general it takes around 0.2 seconds per analysis and redaction of a single document (on a small server around 20,000 words per second), which means that automated anonymization of the original 0.6m document set removing 99% of PII is possible within 1.4 days of server time (instead of 12.5 years) and the ongoing monthly server time for the 10,000 documents is around 33 server minutes.

The end result is that HSE is able to unlock their very valuable data science personnel to drive value towards the *analysis* of the data rather than the *engineering* of the data.

## Next steps

On the technical side, there is further work to be done to further reduce under-redacted PII, with several open issues. With the overall aim of automating the anonymization process, the outstanding issues will be minimised programmatically to the extent possible, for instance by adding extra processing to search the full text a second time, to look for versions of successfully identified entities. Ohalo anticipate that resolving the outstanding issues on a programmatic basis could potentially reduce the 'Significant Breach' count to just 10 records instead of 19, or in other words, 99.5% of records successfully anonymized.

The results of this evaluation are intended to serve as a basis for discussion and to allow a definition of what an acceptable level of risk would be, given the existing controls in place. This will guide future evaluations and identify what the next steps are, either in process improvement, evaluation or data anonymisation.

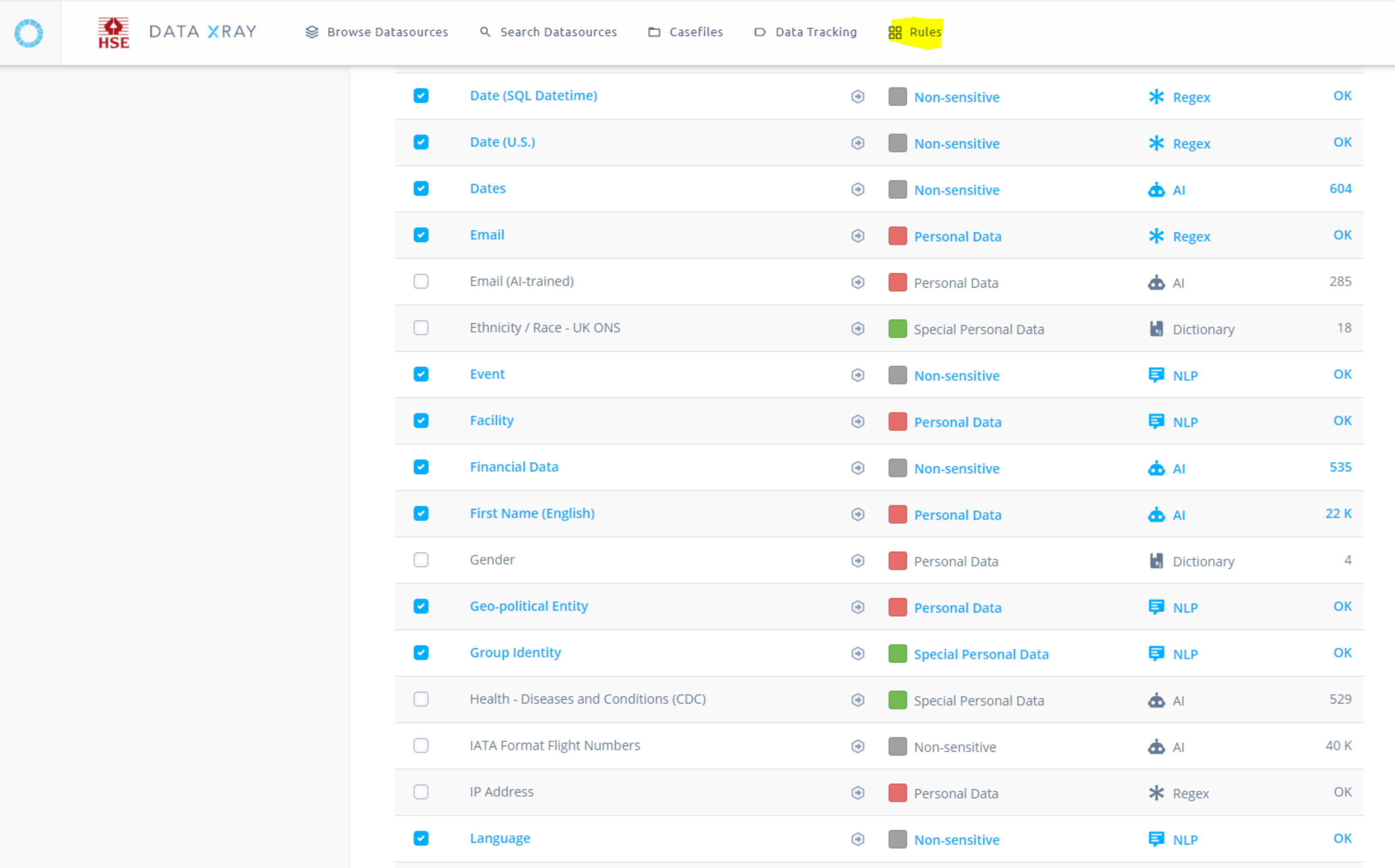
On a practical basis, there are three cases for deploying the Data X-Ray in production for the HSE and their partners. The first is deployment of the Data X-Ray in a production environment to continue the work on anonymization of HSE's current data such as RIDDOR reports.

The second is deployment of a service for HSE's partners where data from HSE's partners can be shared more easily between themselves to prevent death, injury and ill health to those at work and those affected by work activities not only within the UK but hopefully around the world.

Lastly, the third is an HSE project on data anonymisation whose purpose is to anonymise data for long term archival. While still ongoing, the plan is to produce a generic anonymised record for archival and future research when HSE discards the original RIDDOR record (or other data of research interest) after 10 years. Anything like this that HSE generates and has national value will eventually be put in the National Archives.

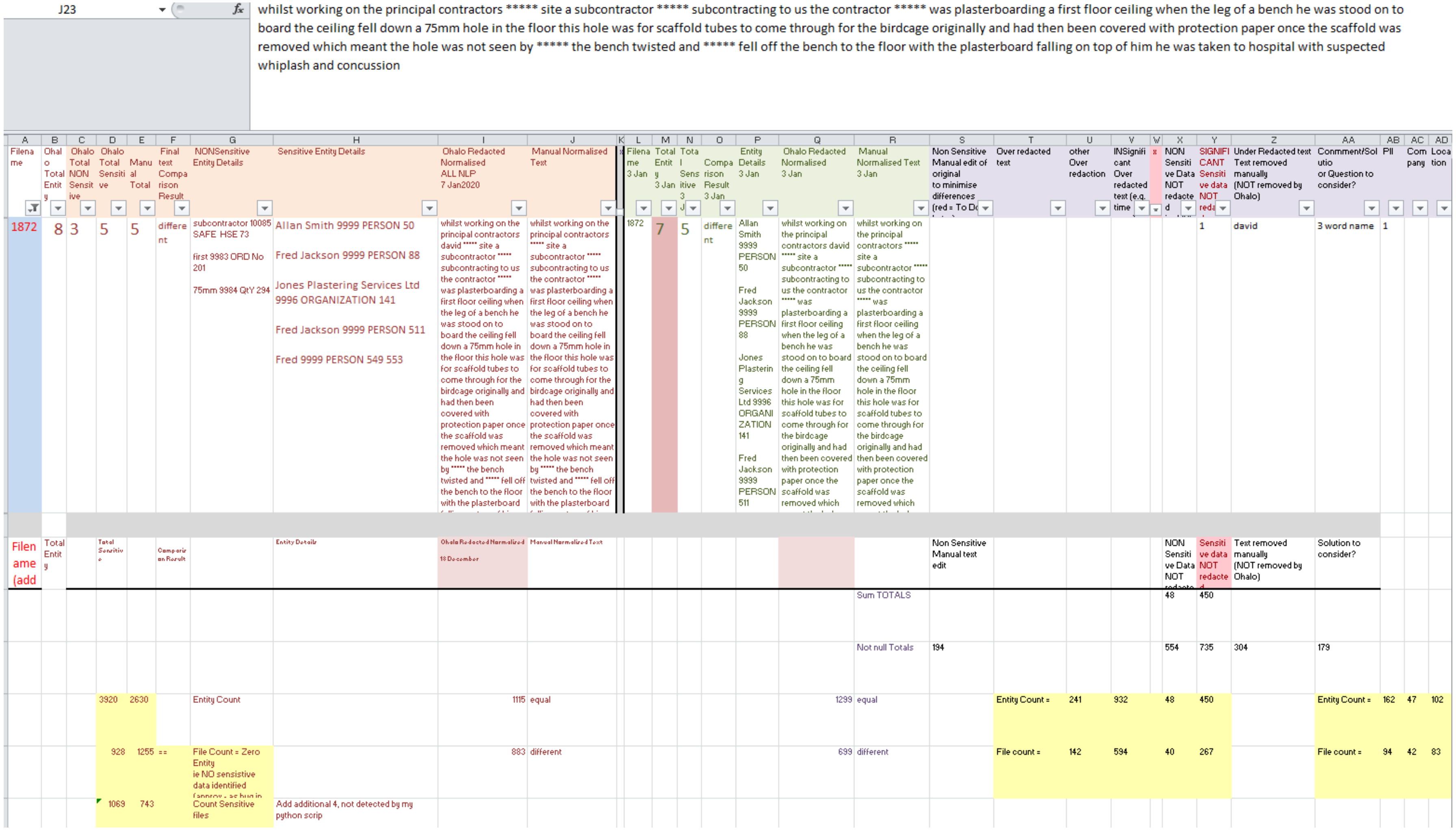
However it is ultimately up to HSE in terms of what an acceptable level of risk is, this is determined by HSE's information asset owners. Assuming that the HSE accepts the risk of 99%+ anonymisation, the intention is to pursue these cases in a production environment.

## Appendix A: Screenshots illustrating evaluation methodology



| | | | | |
|---|---|---|---|---|
| ☑ | Date (SQL Datetime) | Non-sensitive | Regex | OK |
| ☑ | Date (U.S.) | Non-sensitive | Regex | OK |
| ☑ | Dates | Non-sensitive | AI | 604 |
| ☑ | Email | Personal Data | Regex | OK |
| ☐ | Email (AI-trained) | Personal Data | AI | 285 |
| ☐ | Ethnicity / Race - UK ONS | Special Personal Data | Dictionary | 18 |
| ☑ | Event | Non-sensitive | NLP | OK |
| ☑ | Facility | Personal Data | NLP | OK |
| ☑ | Financial Data | Non-sensitive | AI | 535 |
| ☑ | First Name (English) | Personal Data | AI | 22 K |
| ☐ | Gender | Personal Data | Dictionary | 4 |
| ☑ | Geo-political Entity | Personal Data | NLP | OK |
| ☑ | Group Identity | Special Personal Data | NLP | OK |
| ☐ | Health - Diseases and Conditions (CDC) | Special Personal Data | AI | 529 |
| ☐ | IATA Format Flight Numbers | Non-sensitive | AI | 40 K |
| ☐ | IP Address | Personal Data | Regex | OK |
| ☑ | Language | Non-sensitive | NLP | OK |

*Figure 1 : Screenshot of the customizable classifier functionality within Data X-Ray. Classes are redacted in the order : Dictionary, Regular Expression Matching, NLP and lastly Artificial Intelligence.. Users can select which classes to redact, modify existing default classes or create new customised classes.*



J23 fx whilst working on the principal contractors ***** site a subcontractor ***** subcontracting to us the contractor ***** was plasterboarding a first floor ceiling when the leg of a bench he was stood on to board the ceiling fell down a 75mm hole in the floor this hole was for scaffold tubes to come through for the birdcage originally and had then been covered with protection paper once the scaffold was removed which meant the hole was not seen by ***** the bench twisted and ***** fell off the bench to the floor with the plasterboard falling on top of him he was taken to hospital with suspected whiplash and concussion

| A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|
| Filename | Ohalo Total Entity | Ohalo Total NON Sensitive | Ohalo Total Sensitive | Manual Total | Final text Comparison Result | NONSensitive Entity Details | Sensitive Entity Details | Ohalo Redacted Normalised ALL NLP 7 Jan2020 | Manual Normalised Text | x |
| 1872 | 8 | 3 | 5 | 5 | different | subcontractor 10085 SAFE HSE 73<br>first 9983 ORD No 201<br>75mm 9984 QtY 294 | Allan Smith 9999 PERSON 50<br>Fred Jackson 9999 PERSON 88<br>Jones Plastering Services Ltd 9996 ORGANIZATION 141<br>Fred Jackson 9999 PERSON 511<br>Fred 9999 PERSON 549 553 | whilst working on the principal contractors david ***** site a subcontractor ***** subcontracting to us the contractor ***** was plasterboarding a first floor ceiling when the leg of a bench he was stood on to board the ceiling fell down a 75mm hole in the floor this hole was for scaffold tubes to come through for the birdcage originally and had then been covered with protection paper once the scaffold was removed which meant the hole was not seen by ***** the bench twisted and ***** fell off the bench to the floor with the plasterboard | whilst working on the principal contractors ***** site a subcontractor ***** subcontracting to us the contractor ***** was plasterboarding a first floor ceiling when the leg of a bench he was stood on to board the ceiling fell down a 75mm hole in the floor this hole was for scaffold tubes to come through for the birdcage originally and had then been covered with protection paper once the scaffold was removed which meant the hole was not seen by ***** the bench twisted and ***** fell off the bench to the floor with the plasterboard | |
| Filename (add | Total Entity | | Total Sensitive | | Comparison Result | | Entity Details | Ohalo Redacted Normalized 18 December | Manual Normalized Text | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | 3920 | 2630 | | Entity Count | | 1115 | equal | |
| | | | 928 | 1255 | == | File Count = Zero Entity ie NO sensistive data identified (approx - as bug in | | 883 | different | |
| | | | 1069 | 743 | | Count Sensitive files | Add additional 4, not detected by my python scrip | | | |

| L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | AA | AB | AC | AD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Filename 3 Jan | Total Entity 3 Jan | Total Sensitive 3 J | Comparison Result 3 Jan | Entity Details 3 Jan | Ohalo Redacted Normalised 3 Jan | Manual Normalised Text 3 Jan | Non Sensitive Manual edit of original to minimise differences (red = To Dc | Over redacted text | other Over redaction | INSignificant Over redacted text (e.g. time | x | NON Sensitive Data NOT redacted | SIGNIFICANT Sensitive data NOT reda | Under Redacted text Text removed manually (NOT removed by Ohalo) | Comment/Solutio or Question to consider? | PII | Company | Location |
| 1872 | 7 | 5 | different | Allan Smith 9999 PERSON 50<br>Fred Jackson 9999 PERSON 88<br>Jones Plastering Services Ltd 9996 ORGANIZATION 141<br>Fred Jackson 9999 PERSON 511 | whilst working on the principal contractors david ***** site a subcontractor ***** subcontracting to us the contractor ***** was plasterboarding a first floor ceiling when the leg of a bench he was stood on to board the ceiling fell down a 75mm hole in the floor this hole was for scaffold tubes to come through for the birdcage originally and had then been covered with protection paper once the scaffold was removed which | whilst working on the principal contractors ***** site a subcontractor ***** subcontracting to us the contractor ***** was plasterboarding a first floor ceiling when the leg of a bench he was stood on to board the ceiling fell down a 75mm hole in the floor this hole was for scaffold tubes to come through for the birdcage originally and had then been covered with protection paper once the scaffold was removed which | | | | | | | 1 | david | 3 word name | 1 | | |
| | | | | | | | Non Sensitive Manual text edit | | | | | NON Sensitive Data NOT redacte | Sensitive data NOT redacted | Text removed manually (NOT removed by Ohalo) | Solution to consider? | | | |
| | | | | | | Sum TOTALS | | | | | | 48 | 450 | | | | | |
| | | | | | | Not null Totals | 194 | | | | | 554 | 735 | 304 | 179 | | | |
| | | | | | 1299 | equal | | Entity Count = | 241 | 932 | | 48 | 450 | | Entity Count = | 162 | 47 | 102 |
| | | | | | 699 | different | | File count = | 142 | 594 | | 40 | 267 | | File count = | 94 | 42 | 83 |
| | | | | | | | | | | | | | | | | | | |

*Figure 2 : Sample data collated from the python script and manual 'eye-ball' evaluation task (real data has been replaced with contextually appropriate placeholders).*

The orange headed columns were populated by the python script and contain: two versions of the anonymised text (Auto and Manual output), a list the Data X-Ray classified entities (both 'sensitive' and 'non-sensitive'), along with a count of the number of entities extracted from the two versions.

The purple headed columns were populated as a result of manually eye-balling the identified differences; documenting the over and under redacted entities.

The python utility output assists with the comparison of two different versions of the redaction process, for development and regression testing.

## Appendix B: General observed redaction anomalies

### Over redaction

The following regular Over redaction anomalies have been noted:

1. The majority of the over redacted entities relate to phrases associated with periods of time e.g.

| | |
|---|---|
| the next day | DATE |
| the weekend | DATE |

2. Full stop adjacent to an entity or date, often redacts too much (back-end process error)

| | |
|---|---|
| FPS.There | ACRONYM3_HSE |
| point.in plant room | FACILITY |
| 31/03/17.Please | DATE (INTERNATIONAL) |
| 27 May 2013.Given | DATE |

In the above examples the text before or after the full-stop is incorporated into the identified sensitive entity, and is therefore removed without need. This is a low priority issue, but one that would need Ohalo to resolve.

### Under redaction

Having reviewed the types of text that is ‘missed’ by Data X-Ray, there appears to be some categories that repeat regularly, and therefore will be reported and discussed with Ohalo to determine the best resolution solution going forward. This may require some more general modifications to Data X-Ray and/or best rectified through the modification of the customisable ‘Classifier’.

The following regular Under redaction anomalies have been noted (real data has been replaced with contextually appropriate placeholders):

1. Missing Names

   In many of the cases where a name has been ‘missed’, it has also been successfully redacted when positioned in a different location within the same body of text. This anomaly could be minimised programmatically by adding extra processing to search the text again, to look for versions of successfully identified name entities.

   e.g. If the entity “Fred Taylor” is successfully identified, then all instances of Fred or Taylor or FT should also be removed.

2. Double-barrel surname

   Names having a double-barrel surname cause part of the name to be under redacted. The example below resulted in the second part of the surname failing to be redacted.

| | |
|---|---|
| Mr Fred | PERSON |
| Sparkes- | PERSON |