## Version Control

|  |  |  |
| --- | --- | --- |
| Date | Notes | Author |
| 2018 | Initial Draft | Kafico Ltd |

# Exploration: Accuracy and Data Quality

The aim of this section is to explore whether the project aligns with the data protection principle related to the accuracy of personal data.

## Sources

[Data Protection Act 2018 (DPA)](http://www.legislation.gov.uk/ukpga/2018/12/contents)

[General Data Protection Regulations (EU) 2016/679 (GDPR)](https://gdpr-info.eu/art-9-gdpr/)

[Information Commissioner – Guide to the General Data Protection Regulations (ICO Guide)](https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf)

## Definitions / Context

* Controllers must carefully consider any challenges to the accuracy of information
* Controllers must consider whether it is necessary to periodically update the information
* Every reasonable step should be taken to ensure that personal data which are inaccurate are rectified or deleted
* if you are using the data to make decisions that may significantly affect the individual concerned or others, you need to put more effort into ensuring accuracy.

# Assessment

There is not much information from the relevant sources in relation to particular measures for data quality being implemented by Controllers and Processors and it seems logical that it would depend on the nature of the project itself.

There is a requirement for Controllers to ensure that suitable data quality measures are in place including; how users will be trained or instructed to use systems appropriately, how records or electronic transactions will be validated against their source when added to another system, or as a result of direct data entry and how systems will react if transactions or transfers of data are not received properly.

The following is a description of the measures in place to ensure data quality and integrity for the PSL products and services;

**Data Extraction**

PSL have devised an algorithm that identifies when the extracted data set falls outside of expected parameters. Irregularities are highlighted through the presence of unexpected elements i.e. the size of the data set, number of data lines, number of drugs, blood pressure readings. Where the data has characteristics which could be deemed as outliers, the extraction would not be accepted by the system and this would trigger manually scrutiny.

**Data Transfer**

The extracted data is encrypted for transit, in order for the data set to effectively ‘land’, it must decrypt which means that it must be complete. It will only allow decryption and therefore accept the file if the file is complete. The systems have interoperability so rather than show corrupt data, the system will reject it.

**Algorithm Application**

The algorithm is programmed to create alerts when a combination of particular data points are in existence. For example a patient who is on comibination of certain medicines known to react with one another might trigger an alert for a medication review.

The algorithm is programmed using NHS England guidance and is subject to a quarterly clinical review within PSL to ensure that the data upon which the alerts are based remains accurate and best practice.The clinical team within PSL will also undertake periodic audits of alert numbers and other outliers to identify anomonlies – for example, a sudden spike in the number of alerts being issued would trigger a closer look at the data being produced.

Additionally, there is a feedback button available to all end users of the system. This allows users of the system to identify where there might be gaps in the information or perhaps an alert has been innapropriately generated. So, PSL are in receipt of around 10,000 reviews supporting the ongoing development of the service.

**Re-identification**

The system involves a brand new build of the integrated data sets each week. Each build requires the extraction of the data, the replacement of the identifier with the Eclipse integer.

This means that there is low risk of a mismatch between the identifying data (NHS No, Patient Name) and the other extracted items (read codes) when they are pulled back together to facilitate the identification of a particular patient.

There have been no mismatches of this data since the system inception in 2011. The only example where a mismatch between the extracted data and the patient identity would be possible is where the wrong NHS No has been attributed to the patient within the source data and this is outside the scope of control for PSL.

# Conclusion

PSL will provide evidence to Controller customers of validation / accuracy measures undertaken for the following processing activities;

* Data Extraction
* Data Transfer
* Algorithm Application
* De-identification
* Re-identification
* Display / Query

# Risk Assessment

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk / Gap** | **Action** | **Status** | **Responsible Party** |
| There is a risk that PSL do not have suitable data validation protocols in place to ensure effective matching / linkage of data sets | PSL should draft an Accuracy and Data Quality document that identifies that steps are taken for each processing activity that ensure validation and accurate matching / linkage | Complete | PSL |

#