Proactive Validation - Using Anonymised Data for Validation

PURPOSE OF THIS NOTE

This note explains:

- Part A: What is Proactive Validation?
- Part B: Why can employers send Salary Finance data proactively?

PART A: WHAT IS PROACTIVE VALIDATION?

- When an employee starts an application for a product (such as a loan) on the Salary Finance website, they will be asked to input the following four data points: payroll ID, date of birth, salary and employment start date.

- SF needs to check that this data is accurate by comparing it with the data the employer holds. Salary Finance therefore asks employers, on implementation and at agreed cadences after that, to use Salary Finance's Employee Validation Application (EVA) to completely anonymise a copy of its employee database before sending it to SF. SF stores the anonymised data and when an employee applies for a product, Salary Finance can quickly check that the data the employee provides is accurate.

PART B: WHY CAN EMPLOYERS SEND SALARY FINANCE DATA PROACTIVELY?

Encryption and hashing aspects of SF's Employee Validation Application (EVA) - Outline

- Salary Finance does not need - and does not want to - obtain or hold the actual content of the employer database. It only needs to check that there is a match with the data Salary Finance receives from employees who apply for products on its platform.

- This can be achieved by using data that has been depersonalised. EVA encrypts and hashes the employer database before the database is uploaded to the Salary Finance platform (in the case of Proactive validation). The Salary Finance platform matches the encrypted and hashed data against the encrypted and hashed data that it holds about employee applicants.

- Neither Salary Finance nor anyone else presented with the encrypted and hashed data could realistically reverse the hashing process and unencrypt the data.

Encrypted and hashed data provided by the employer is not Personal Data

- Salary Finance has sought legal advice, including a legal opinion from a barrister at 11 King's Bench Walk. The barrister noted that the "*pivotal question for present purposes was whether that hashed data [i.e. the data provided by the employer] constitutes personal data for GDPR purposes*".

- He concluded that the answer is no and that Salary Finance is "*disclosing anonymised information that does not constitute personal data...When the employer sends the hashed dataset to Salary Finance, it is not disclosing personal data, and Salary Finance is not receiving or using personal data*."

The reasons are as follows:

- Personal Data is defined in Art 4(1) of GDPR as "...information relating to an identified or identifiable person…".

- Whether data relates to an identifiable person must be approached in accordance with Recital 26.

- Salary Finance recognises that Recital 26 states that "*Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person*".

- However, it is not the case that all personal data that has undergone pseudonymisation remains to be Personal Data. This is clear from the GDPR and case law for two reasons: (1) identification of a person from the data must be reasonably likely; and (2) there is a clear distinction in the GDPR between anonymous and pseudonymised data. Considering each in turn:

(1) The 'reasonably likely' test

- It is clear from Recital 26 that it is only pseudonymised data that "could" be attributed to an identified person that remains to be Personal Data i.e. not simply all pseudonymised data.

- Recital 26 explains when data "could" be attributed to a natural person. It states:

> "*... To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments…*"

- Whether identification of a person is "reasonably likely" has been considered in the following case law:

  - *Department of Health v Information Commissioner [2011] EWHC 1430*: Low cell count statistics on abortions were not personal data, as the possibility of a third party identifying any individuals from those statistics was "*extremely remote*".

  - *Breyer v Bundesrepublik Deutschland (Case C582/14) (Court of Justice of the European Union)*: That case concerned a German public authority that retained the dynamic IP addresses of visitors to its website. It was held that the theoretical possibility of an internet service provider unlocking the identities of visitors was not enough to render the anonymous IP addresses personal data:

    "*... it must be determined whether the possibility to combine a dynamic IP address with the additional data held by the internet service provider constitutes a means likely reasonably to be used to identify the data subject.*

    ... that would not be the case if the identification of the data subject was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant."

- Identification of individuals from the encrypted and hashed data from the employer is not reasonably likely. No person can realistically reverse the hashing process and unencrypt the data.

(2) The distinction between pseudonymised data and anonymous data

- Recital 26 makes a clear distinction between pseudonymised data and anonymous data:

  "*...The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the <u>data subject is not or no longer identifiable</u>. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.*"

- This is further supported by the latest ICO Guidance which provides that "*if data can be truly anonymised, then the anonymised data is not subject to GDPR*".

- Furthermore, it is supported by the ICO's Code of Practice on Anonymisation (issued prior to the introduction of the GDPR but remains useful guidance). This states:

> “*There is clear legal authority for the view that where an organisation converts personal data into an anonymised form and discloses it, this will not amount to a disclosure of personal data. This is the case even though the organisation disclosing the data still holds the other data that would allow re-identification to take place. This means that the DPA no longer applies to the disclosed data*.”

- Where data has been pseudonymised such that identification of a person from such data is no longer reasonably likely, such data is not Personal Data and the GDPR does not apply.