



Ministry
of Justice

Disclosure Team
Ministry of Justice
102 Petty France
London
SW1H 9AJ

data.access@justice.gov.uk

MoJ ref: 260715061

10 September 2026

Dear Mr [REDACTED]

FREEDOM OF INFORMATION ACT (FOIA) INTERNAL REVIEW

Thank you for your Internal Review request dated 15 July 2026 regarding FOI request 260616044 in which you asked for the following information from the Ministry of Justice (MoJ):

I am writing to request recorded information under the Freedom of Information Act 2000. This request concerns the Ministry of Justice's use of the open-source probabilistic record-linkage library Splink, as described in the MoJ's published Algorithmic Transparency Recording Standard record "MoJ: Splink Master Record." That record states that Splink is used operationally — including a real-time Core Person Record (CPR), a Probation in Court function, and a pilot sharing Police National Computer numbers with police — to link individuals across HMCTS court, HMPPS prison, and probation case-management systems (DELIUS, NOMIS, Common Platform, LIBRA, FamilyMan, CaseMan).

I request the following recorded information:

- 1. Governance documents. The complete Data Protection Impact Assessment(s) and the internal Algorithmic Transparency Record(s) for (a) the Core Person Record service and (b) the Probation in Court system, in full or with only those redactions strictly necessary under the Act (please itemise any exemption applied, with the public-interest balancing where engaged).**
- 2. Model configuration. The Splink settings/configuration for these systems (JSON settings dictionaries and/or Python model scripts), specifically including: the blocking rules used to partition records; and the comparison definitions / comparison vectors, i.e. the comparison levels and string-similarity thresholds (e.g. Levenshtein, Jaro–Winkler, or Jaccard cut-offs) applied to forename, surname, date of birth, and address fields.**
- 3. Decision threshold. The match-weight (log2 Bayes factor) threshold, or equivalent posterior match-probability cut-off, at or above which two records are treated as a definitive link within operational court/probation workflows; and any separate "clerical review" band between auto-accept and auto-reject thresholds.**
- 4. Error performance. The target and observed false-positive and false-negative (false-match and missed-match) rates accepted for these operational systems; and the methodology for estimating the m and u probabilities for each matching parameter (including use of the Expectation–Maximization algorithm and any labelled/clerical ground-truth set used to validate accuracy).**
- 5. Human review & redress. Recorded information on whether, and at what point, a human reviews a candidate match before it takes operational effect (including before each PNC share in the probation → police pilot); and the process by which an individual affected by an erroneous link can have it corrected.**

The purpose of an Internal Review is to assess how your FOI request was handled in the first instance and to determine whether the original decision given to you was correct. This is an independent review. The reviewer of this case was not involved in the original decision.

The response to your original request confirmed the information requested is held and the MoJ confirmed that it does not hold separate internal Algorithmic Transparency Records for the Core Person Record (CPR) or Probation in Court services. A section 21 exemption was applied to Splink Master Record and a link to this was provided as it was deemed that this information is already reasonably accessible to you. The MoJ also confirmed it holds a DPIA for CPR but withheld it under section 31 of the FOIA on security and crime prevention grounds, with some information exempted under section 40(2) as it was deemed to be third-party personal data. For model configuration, matching thresholds, error methodology and human review processes, you were directed to publicly available source code and documentation on github. MoJ did not provide further detail under exemption section 21 of the FOIA. After careful consideration we have concluded that this response was partially compliant with the requirements of the FOIA.

Statutory deadline

The statutory deadline for your request was 16 June 2026, and the response was provided on 15 July 2026 which included an extension period in order to conduct a public interest test. Therefore, the response was compliant with the timeliness requirements of the FOIA.

Outcome

1. Governance documents: DPIA and Algorithmic Transparency Records

The business area has confirmed that the Ministry of Justice does not hold a separate internal Algorithmic Transparency Record for the services in scope. Any relevant information held is within the published record. I therefore uphold the application of section 21. I have provided the link again, for ease.

The published record can be accessed at: <https://www.gov.uk/algorithmic-transparency-records/moj-splink-master-record>

I have reconsidered whether section 31(1)(a) justifies withholding the Data Protection Impact Assessment (DPIA) in full and whether information can be separated and disclosed with redactions. I have concluded that section 31(1)(a) applies to parts of the DPIA where disclosure would be likely to prejudice the prevention or detection of crime by revealing detailed matching methodology, linking and unlinking processes, system architecture or data flows, security controls and assurance information, or operational law-enforcement datasets and identifiers. However, the exemption does not justify withholding the document in full. I have therefore enclosed a redacted copy of the DPIA. In balancing the public interest, I have given weight to transparency and accountability concerning the use and governance of data-matching technology, whilst recognising the public interest in protecting information that could expose operational systems or controls to misuse. I consider that the balance favours disclosure of the non-sensitive material and withholding only the identified sensitive passages.

2. Model configuration

The service team has confirmed that the published repository contains the requested information about how the model is configured, including the model files relevant to blocking and comparison configuration. I therefore uphold the application of section 21 to this information as the link has been provided to an open source.

The information can be accessed at: https://github.com/ministryofjustice/hmpps-person-match/tree/main/hmpps_cpr_splink/cpr_splink/model/model_files

This finding relates to the published model configuration. The service team has separately noted that the repository does not disclose operational data about matching.

3. Decision threshold

The original response explained that operational matching thresholds are defined in published source code and stated that there is no separate clerical-review threshold when records are initially matched. It further explained that an existing link may enter a clerical-review queue where updated source data reduces the score below the fracture threshold of 18.

The source code can be accessed at: https://github.com/ministryofjustice/hmpps-person-match/blob/main/hmpps_cpr_splink/cpr_splink/model/model.py

There are no separate unpublished thresholds.

4. Error performance

I uphold your complaint on this item. The original response directed you to model configuration used during training, but that did not directly answer the request for target and observed false-positive and false-negative rates.

Following further enquiries, the business area has confirmed that the MoJ does not hold recorded information falling within this part of the request. No target or accepted false-positive or false-negative rate has been set for these services. The correct response under FOIA is therefore that this information is not held.

The FOIA does not oblige a public authority to create information to answer a request if the requested information is not held. The duty is to only provide the recorded information held.

For context, the business area has explained that a population false-positive or false-negative rate would require an independently verified ground-truth dataset showing which records across the source systems genuinely relate to the same person. It has confirmed that no such reference set exists for the prison and probation population.

The business area has also explained that assurance is instead applied through conservative match-weight thresholds, human review during model development of record pairs close to the threshold, clerical review of sampled pairs, and the ability to break an incorrect link and record a manual override. Clerical review of sampled pairs may indicate disagreement with links made by the model, but it does not establish a population error rate or identify all missed matches.

We can confirm that the MoJ holds information relating to the methodology used to estimate the m and u probabilities used during the training of the Core Person Record matching model. Please see attached document (The methodology used for estimating m and u probabilities)

5. Human review and redress

The published Algorithmic Transparency Record contains information about human review, and I uphold section 21 for information reproduced in that record.

The business area has clarified that candidate matches at or above the joining threshold are linked automatically and are not individually reviewed by a person before the link is created. Human involvement occurs when linked information is presented to a member of staff in an operational service, such as Probation in Court, where it is used as supporting information alongside other available information. Where an incorrect link is identified, it can be corrected and a manual override recorded so that the same link is not remade. The original statement that "Splink does not directly make decisions about individuals"

was too absolute. The clearer position is that Splink produces a probability that two records describe the same person and records are linked where the applicable threshold is met. Decisions about individuals are made by people using the information available to them, of which a linked record may form one part.

Where an incorrect link is identified it can be corrected in the system, and a manual override recorded so the same link is not remade. Individuals who believe information held about them is inaccurate may contact the relevant justice service responsible for the record in question and request that the information be reviewed and corrected where appropriate.

<https://www.gov.uk/algorithmic-transparency-records/moj-splink-master-record#human-review>

Appeal Rights

If you are not satisfied with this response, you have the right to apply to the Information Commissioner's Office (ICO). The Commissioner is an independent regulator who has the power to direct us to respond to your request differently, if they consider that we have handled it incorrectly.

You can contact the ICO at the following address:

<https://ico.org.uk/make-a-complaint/foi-and-eir-complaints/>

Postal address

Information Commissioner's Office
Wycliffe House
Water Lane
Wilmslow
Cheshire
SK9 5AF

Yours sincerely

Justice Digital, Data and Science