Ministry of Justice – Reducing Reoffending Intern

TIN-MOJ-001

About the Organisation

The Ministry of Justice (MoJ) is a ministerial government department responsible for running over 100 prisons, 500 courts in England and Wales, as well as probation services. The department’s priorities are to protect the public from serious offenders, to improve the safety and security of our prisons, reduce reoffending, and deliver swift access to justice.

The MoJ works with many other government agencies including HM Prison and Probation Service (HMPPS) and Youth Justice Board (YJB). The department conducts an extensive programme of analytical and research-based work, helping to deliver a world class justice system. Diverse teams of experts from different fields work together to conduct research reviews, commission research, provide justice statistics and secondary data analysis, cost benefit analysis, forecasting, advanced analytics and data engineering.

Role Description and Responsibilities

Context – BOLD Reducing Reoffending pilot

The intern will be part of the BOLD programme – Better Outcomes through Linked Data, which was set up in April 2021 to improve the connectedness of government data in England and Wales. Government is responsible for supporting people with multiple complex needs such as those who are homeless, victims of crime and offenders, however this is difficult because information about individuals is held across different government departments and not frequently shared.

The intern will work with the Reducing Reoffending pilot team within the BOLD programme. Reoffending costs society approximately £18 billion a year. By linking data to understand the impact of interventions relating to employment, health, housing and family, the pilot team hope to show how data and evidence can be used effectively to better understand how to help offenders turn their backs on crime.

Families work strand

The specific work strand will involve working with the Reducing Reoffending pilot team on the “Families and Intergenerational Offending project”. The intern would develop a reproducible analytical pipeline extracting information held digitally in free text across multiple HMPPS sources.

The ambition of the Families project is to develop an annual estimate of the number of children of prisoners with national and subnational breakdowns. Initially this data will come from the
internal data sources previously stated, with data from other government departments and agencies to be linked in at a later stage.

**Analysis of free text data**

HMPPS data is vast and complex reflecting the complexity of real-life prison and probation processes. Much of the data required for this project is stored in unstructured form in long, complex sentences, or in note form and with abbreviations and acronyms. There are typically thousands of notes per offender, and probation and prison officers all have unique writing styles and approaches.

The intern would use techniques such as natural language processing to accurately extract information to help build estimates of the number of children of prisoners with associated geographical and demographic breakdowns. Any model developed would need to be trained on HMPPS text and robustly validated to provide an indication of the model’s confidence. This is a challenging project as the intern would need to spend some time understanding peculiarities in the free text. The goal would be to work with the Families project team to create a reproducible analytical pipeline linking the extracted data with other relevant (non-free-text data) to generate annual statistics. The work should include an evaluation of the accuracy of this pipeline.

**Expected Outcomes**

Deliverable 1 – Develop code to extract information on the offender’s family from HMPPS data sources. Evaluate the model confidence in order to establish reliability of the information extracted.

Deliverable 2 – Consider the possibility of developing a RAP to link the information extracted, combining with other data (in non-free-text fields) already collected by the team, to develop annual statistics on children of prisoners/ prisoners with children. It would be possible to use R, Python, Athena, AirFlow, DBT. The Families project team should be involved in all steps to build and automate the pipeline. It is essential that there will be clear and detailed documentation, communication and hands-on trials from the early stages of the process. If this is not possible then documentation should be produced on the limitations (for example with reference to hosting, interfaces, APIs, format constraints, security or other factors) and how these could potentially be overcome.

Please note that data held across the prison and probation administrative systems is vast and held across multiple tables which can be linked. There would be initial period for the intern to develop their understanding of the data. There is the possibility for the scope of this project to be extended to include data from other government departments.

**Impact of this work**

Parental imprisonment is acknowledged as an Adverse Childhood Experience (ACE). Longitudinal studies in the UK and Europe have strongly linked parental imprisonment to negative outcomes for children including complex emotional needs, low levels of engagement in education, future offending behaviour.

As part of the Prison Strategy White Paper (2021), the MoJ has made commitments to investigate data sharing on offenders and their families and to understand the number of
children affected. The Farmer Reviews (2017 & 2019) highlighted issues of intergenerational offending and identified that there is no official process for recording the number of children impacted. The Joint Committee on Human Rights: Children of Mothers in Prison & the Right to Family Life (2021) as well as the Justice Select Committee (JSC): Women in Prison (2022) both highlighted the lack of official data on maternal imprisonment. The government response to JSC committed to improving data sharing and understanding through BOLD.

The overall impact of this work is to meet the commitments set out above, and to create evidence which can inform effective policy intervention, specifically relating to children of prisoners, and their families.

Supervision and Mentorship

Joint supervision will be provided by the team at MoJ and the Head of AI for Public Services at the Alan Turing Institute.

Dr Caroline Tudor (MoJ BOLD programme – lead analyst for the Reducing Reoffending pilot)

Dr Andreas Soteriades (MoJ BOLD programme – lead data scientist for the Reducing Reoffending pilot)

Dr Jonathan Bright (Alan Turing Institute - Head of AI for Public Services)

Caroline is the MoJ project analytical lead and will provide direction on the HMPPS data and how the model should be implemented.

Andreas will provide some additional data science oversight from an MoJ perspective to ensure the pipeline developed can be taken forward after the duration of the project.

Jonathan will supervise the AI/data science work of the intern with fortnightly remote meetings.

Person Specification

Essential expertise and skills:

- Understanding of working with sensitive and confidential datasets.
- Experience researching within the NLP domain, especially with regard to knowledge extraction, summarization, or NER.
- Experience using R, Python, sql, git.
- Experience modularising code, writing functions, using functional programming, and automation.
- Proven track record in research software development.
- Good interpersonal skills and ability to build and maintain strong working relationships with partners, stakeholders and fellow analysts.
- Ability to take a complex problem and develop analytical solutions to inform decision-making.
- Strong ability to communicate key messages in writing and verbally to a variety of audiences.

Preferred knowledge and experience:
- Any knowledge of the criminal justice system would be a significant advantage but not required.
- Desirable AirFlow, DBT
- Data integration and MLOps expertise, e.g. Databricks/AzureML/AWS SageMaker
- Experience contributing to large-scale open source or other code projects or conducting code review
- Experience writing documentation at several levels of assumed technical competence
- Familiarity with civil service ways of working for example awareness of the Code of Practice.

**Internship Logistics**

**Start Date:** September 2023  
**Salary:** Starting from a minimum of £31,265 per annum pro rata  
**Location:** London or national hubs

**Security Clearance:** Basic security clearance required as well as training and signing of internal agreements with respect to each data source.

**Internship Duration:** The intern would work full-time for 3-6 months (a part-time option for 6-12 months may be considered).  
The intern would join as a civil servant. MoJ policy is to attend the headquarters in London or national hub twice a week. Fully remote working is also an option.

There are strict nationality rules for civil servants. Only the nationals represented in this list can be employed as civil servants:

- UK nationals
- Nationals of Commonwealth countries who have the right to work in the UK
- Nationals of the Republic of Ireland
- Nationals from the EU, EEA or Switzerland with settled or pre-settled status or who apply for either status by the deadline of the European Union Settlement Scheme
- Relevant EU, EEA, Swiss or Turkish nationals working in the Civil Service
- Relevant EU, EEA Swiss or Turkish nationals who have built up the right to work in the Civil Service
- Certain family members of the relevant EU, EEA, Swiss or Turkish nationals