Sai
By email: request-969704-697be9ac@whatdotheyknow.com

01 June 2023

Ref: **11263**

Dear Sai

Freedom of Information Act 2000 Request

I refer to your Freedom of Information (FOI) request which we received on 13 April 2023 and, more recently, your email of 08 May where you provided more information to enable us to confirm the validity of you identification. I regret I find no record of the CPS receiving your email of 13 April.

I am sorry your initial request was not treated as valid. Further guidance regarding the use of mononyms has now been provided to the Information Access Team (IAT).

The FOI Act gives you the right to know whether we hold the information you want and to have it communicated to you, subject to any exemptions which may apply. It is a public disclosure regime, not a private regime. This means that any information disclosed under the FOI Act, by definition, becomes available to the wider public.

Request

I am interested in knowing the patterns of prosecution of all the various offences prosecuted by CPS; the effects of different sentencing decisions on recidivism; differences in sentencing for pleas vs convictions; changes over time; differences in prosecution rates and outcomes by locality; rates of successful prosecutions depending on various factors; etc.

My research interests include cross-linking and complex analyses such as "recidivism by charge and sentence, controlled for all other factors", "ultimate outcome by initial charge", "sentence severity by post code and ethnicity, controlled for charge severity", "prosecutorial/judicial decisions by economic efficiency / net societal benefit", "likelihood of pleading guilty by pre conviction detention duration, controlled for likelihood of conviction at trial based on all other factors", "primary (eigenvector/regression) factors for likelihood of plea/conviction/sentence/recidivism", "categories of LiP cases by greatest difference between LiP conviction rate/sentence vs likely conviction/sentence if represented by counsel", "effectiveness of prosecution/defence counsel controlled for difficulty of case", "effective causes of general deterrence / specific deterrence / rehabilitation / restoration", etc.

Basic approach of this request

I would like to obtain — in a later request — the raw data necessary to independently make such analyses. They are broadly similar to, but superset of and not covered by, the many statistical publications by the Government and Judiciary relating to offences.

I am first making this request under the FOIA with a view to understanding what records you have that would be most relevant, obtainable, and useful to my research — and in what formats, with what searchability / redactability / aggregablity, with what redundancy with data already publicly available, etc. — in order to later make a separate FOI request for the actual data.

Intended database to be compiled by me

Ideally, as an outcome of my later request — including requests to other entities for related or linkable data — I would like to be able to take data from you and others and use it to synthesise, by my own additional efforts, a fully structured database amenable to analysis which can answer my research questions above (among others).

My intention is to make my own resulting work freely available to the public.

At present, I expect I would need something roughly like the following:

* for each case brought by CPS (i.e. ever seen by a court or defendant, regardless of later changes), to the entire extent records exist in reasonably accessible form, excluding cases that are pending:
- the associated case ID numbers
- the date first brought
- the CPS organisational subdivision responsible (e.g. by locality, specialty, etc)
- the firm, name & bar number of each prosecution and defence counsel
- the trial and sentencing court(s), and names of judicial officer(s), before whom it was brought (these may be different in the case of e.g. a magistrates' court holding trial but moving the case to crown court for sentencing)
- for each defendant:
-- non personally identifying demographic details (e.g. their age, gender, ethnicity, immigration status, first part of post code, etc. at time of offence) and any other statistically aggregable details (e.g. not an English speaker, drug addiction, mental health issues, homeless, etc)
-- some kind of uniquely identifying information (like a hashed identification number) that would enable connecting different cases involving the same defendant while minimising personally identifying data (like name and address) which may be subject to restrictions or should be abonymised due to a sentence being "spent"
- the full list of specific charges (i.e. the most specific formal definition, like a sub-sub-paragraph of a particular act, or particular common law offence, as it would appear in a charging document), including any offense whatsoever (e.g. fines or other minor cases, summary only, either way, indictment only), and for each charge:
-- if not in the first charging document, the date first brought in this case (e.g. a superseding indictment)

-- the list of aggravating or mitigating factors alleged which are amenable to categorisation or discretisation (i.e. primarily any category or list entry set forth in the Equal Treatment Bench Book or Sentencing Council Guidelines' sections on culpability, harm/severity, aggravating/mitigating factors list, etc.), preferably using a standardised code such as exact statutory reference, culpability × harm category (e.g. "A2"), etc.
-- the sentence initially sought by prosecution (including whether concurrent or consecutive, etc.)
-- the post code in which the offence was alleged to have occurred
-- the dates of trial and of sentencing (if any)
-- the outcome of this charge (e.g. judgment in default, dropped due to plea, pled guilty, pled not guilty found not guilty, pled not guilty found guilty, pled not guilty found no case to answer [e.g. on motion to dismiss mid or pre trial], nolle prosequi, Mental Health Act committal, deportation, etc.)
-- the culpability/harm category, aggravating/mitigating factors, etc. (as above), as determined by the court
-- the actual sentence or other terms imposed by court (including e.g. bail amount and conditions; deferred or suspended sentence; early plea discount; sentencing conditions; custodial sentence duration and terms; various forms of discharge and binding over; various forms of community service, ASBO/CBO, restitution, victim compensation, programme participation, and other non custodial sentences; etc.)
-- the actual sentence served on this charge (e.g. due to early release on parole, for good behaviour, etc.)
-- if initial outcomes involved something whose outcome could only be known later (e.g. conditional bail, deferred or suspended sentence, rehabilitation programmes, etc), the ultimate outcome of that (e.g. remand to custody for violation of conditions, successful or non successful completion of rehabilitation programme, finishing suspended period without further offences and therefore sentence being withdrawn, offence being spent, etc)
- whether this case was appealed, and if yes, with what outcome
- the gross costs incurred in prosecuting this case (separate from HMCTS or Judiciary costs), and any offsets that change the net cost of prosecution (e.g. as a result of orders requiring the defendant to pay costs or restitution) … etc.

 # Considerations as to your data

I expect that the information I seek is likely to not be found in a single database, but rather scattered across several different ones, with some codes by which they can be linked. For instance, a lawyer on a case may be linked by bar number, with a separate database of firm affiliation; a charge may be listed by code, which needs to be cross referenced in a lookup table; an appeal may be in a different database, linked by a reference to the lower court case ID; recidivism would likely need to be found by linking cases via a defendant identification number; rehabilitation programme participation and outcomes may be linked to an inmate number rather than an identifier that can link different cases; etc.

I also expect that some of this information may be only found in records that are not amenable or available to statistical analysis — e.g. paper, sealed, privileged, etc. As my interest is statistical analysis, and I do not wish to contest claims of seal or privilege, these would not be of interest to me.

I would prefer to minimise any data which is or may be subject to a right to privacy, such as the identity of defendants whose conviction has been or may yet be spent, or identity of victims — but will need at least some way to link separate cases involving the same person.

 # Request
I respectfully request that you give me a full description of and technical information about all bulk data sources ("relevant data") that
a) are reasonably likely to assist me in compiling, by my own additional efforts, any part of my stated database plan, or
b) are otherwise reasonably relevant to my stated research interests.

This includes any such information available or known to you, regardless of whether you would hold the relevant data. I will follow up separately with whoever does hold it, and would like to have all information that would assist me in doing so. I understand and accept that your information about others' data is not definitive and may be inaccurate — and I request it anyway, as I believe it would be informative.

This also includes any such information about relevant data where the data is already publicly available — both as signposting, and because you likely have additional unpublished information which would make the publicly available data substantially more useful.

 # Limitations and clarifications

Please note that this request does not ask for the actual content of any data sources (but, as below, does cover manuals, technical documentation, and similar descriptive information). In this request, I simply want to understand what data is available — with particularity enough to assess its utility for my research, how to limit your costs, how I may need to integrate multiple data sources, etc. In particular, my intent now is *not* to request the underlying raw data; I will request it in future after first understanding what is available and any considerations that would best optimise the effectiveness of that request.

Consistent with the second paragraph of my request, I specifically request that you not delegate this request to other entities if doing so would cause your costs of response to exceed the reasonable limit. If the relevant data is not held by you, please limit your response to your own knowledge or records — though this includes your knowledge/records about others' data.

I particularly request that your response cover, to the best of your knowledge and records you hold,

A. any relevant data which is stored in a bulk-processable database format, including technical formats such as SQL, JSON, CSV, etc and statistical or research formats such as created by R, MATLAB, etc.;

B. any relevant data which was compiled for any similar research (e.g. the data actually used as input by the researchers, not just the derivatives thereof which are the outcomes of that research);

C. details of relevant data's database/spreadsheet format, structure, enums, query functionality (for FOIA purposes), content, data quality, coverage, limitations, etc., to assist me in understanding what is obtainable within the reasonable costs limit that would best serve my research interests;

and conditionally:

D. if you know of any other entities that hold relevant data for the purposes of the FOIA (even if you also hold the data):
i. the identity of the entity
ii. what data you believe that entity holds iii. what you believe that entity names that data, where you believe it would likely be found or filed, and any other information you have that would aid me in requesting (or aid that entity in finding) that data

E. if you do hold relevant data for the purposes of the FOIA, or if you would be a source of any exemption (even if held by another entity):
a description of any concerns or blanket exemptions which would likely be raised against release under the FOIA, including what parts of what data would be affected to what degree (considering e.g. the options of partial release, release only in certain aggregates, release with randomly altered or hashed data so as to protect privacy without altering statistical analyses, etc);

N.b. on this item: I of course do not expect you to commit to any particular decision as to a future request. Rather, I am asking you to assist me in anticipating and understanding any concerns likely to be arise, and possible mitigations to allay them.

F. to the extent it would stay within the reasonable costs limit while answering the above:
i. all meta documentation — technical documentation, manuals, data quality assessments, usage notes, data schema, or the like — about the relevant data, particularly documents that are primarily used by researchers, statisticians, or database administrators ii. if providing the meta documentation requested in the preceding paragraph would exceed the reasonable costs limit, then instead please provide as much information about such meta documentation as you can while staying within the costs limit — e.g. document titles, authoring bodies/teams, linked data sets, number of pages/bytes, format, date/revision, location within your system of records, and the like

 # Form and format

If a responsive record is already publicly available online, I request that you please provide a direct link to it and reference to the section (e.g. paragraph #, page #, text snippet) that addresses this.

If it is not, I request that you please provide it in its original, native, electronic format, as stored on your computer system.

 # Likely sources of responsive information

I appreciate that this request likely appears complex. However, I do not here actually request a large number of records, nor from a large number of teams.

My expectation is that you maintain only a few databases with relevant data, each of which will have some schema and descriptions as well as a handful of meta documents. I likewise expect that all information I request here is held by only a couple teams — namely, your research/statistics team, and your IT/database management team. They too are the ones I believe most likely to know about other sources of relevant data.

I expect that other researchers, both internal and external, will have conducted closely related analyses before — and therefore that you likely already have all the information requested already collected. What I requested is what any researcher or statistician would need to know — for properly citing, differentiating, or analysing prior work; assessing whether they can answer a question put to them; discovering new research possibilities; formulating a research plan and writing the sources, methods, & background sections of a publication; merging and normalising raw data into a useful format; compiling and analysing their work; and obtaining peer review.

In short, it is just the basic necessities for anyone beginning a research project in this area.

 # Closing

As I do not know what relevant data is potentially available, my request is stated broadly. This is primarily a request for your advice and assistance — in making a future FOI request, and in understanding data sets that would be most beneficial to my research. As this is a core public function under s. 16 FOIA, I trust that you will interpret it accordingly.

If responding to this request would exceed your reasonable costs limit despite all the limitations above, please provide a specific proposal for a modification of my request which would not exceed the limit while being most beneficial to my stated objectives (to the best of your judgment) — i.e. if you would otherwise give a denial on the basis of costs, please instead give me a clarifying response to which I can simply say "yes, I agree with your proposed revision" without having to guess at what is or isn't reasonably feasible for you.

Response

As you stated in your email your request is for "*…primarily… for your advice and assistance*", I respond as such.

The CPS does hold data in relation to your request.

As the CPS deals with material that is sensitive and personal to individuals the underlying raw data is not accessible by the general public.

Applicable exemptions would be, for example: -

- The majority of the data held would likely be exempted under Section **40(2) – Personal Data Relating to Third Parties** or other linked personal data exemptions. For example, the primary ID number is the Unique Reference Number (URN) which is created by the police. The URN is the personal data of the defendant. Such exemptions would also likely apply to any other personal

data held, such as sentencing details (please note here, that, unlike other countries such as the United States, sentences are wholly the responsibility of the sentencing court and are not "sought" by the prosecution), ages, post codes, gender, and ethnicity.

- **Section 12(1) and (2) – cost limit** either to establish specific data if that data is known to be held or to establish whether the data is held at all. In 2021/2022 the CPS brought 426,000 prosecutions and to establish, for example the date of first hearing, each case would have to be examined.

- **Section 31(1)(a) – Law Enforcement –** where providing full details of providers and specific infrastructure hardware models in use for our business-critical systems would prejudice the effectiveness of the systems in place to prevent cyber-attacks relating to internal computer systems and network.

Some data is not held, for example the CPS does not record time spent by internal lawyers, paralegals, and administrative staff on a case-by-case basis. Prosecuting Counsel costs are therefore only available but may be exempted dependent on the nature of the request.

The CPS does publish statistical data such as 'Case Outcomes by Principal Offence Category' on its public website at https://www.cps.gov.uk/publication/performance-management-and-case-outcomes. This data is published in CSV (comma-separated values) format.

The official statistics relating to crime and policing are maintained by the Home Office and the official statistics relating to sentencing, criminal court proceedings, offenders brought to justice, the courts and the judiciary are maintained by the Ministry of Justice. Details are found at https://www.gov.uk/government/organisations/ministry-of-justice/about/statistics.

If you have a research proposal and wish to collaborate with the CPS then guidance on so doing is found at https://www.cps.gov.uk/publication/areas-research-interest.

If you are not satisfied with this response, you have the right to request an internal review by responding in writing to the address below within two months of the date of this response. The internal review will be handled by a member of CPS staff who has not been involved with your original request.

IAT@cps.gov.uk

Information Access Team
Floor 8
102 Petty France
London
SW1H 9AJ

You do have the right to ask the Information Commissioner's Office (ICO) to investigate any aspect of your complaint. However, please note that the ICO is likely to expect internal complaints procedures to have been exhausted before beginning their investigation.

The Information Commissioner can be contacted at:

Information Commissioner's Office
Wycliffe House
Water Lane
Wilmslow
Cheshire **SK9 5AF**
icocasework@ico.org.uk

Yours sincerely

J Bullimore
Information Access Team
0203 357 0788
IAT@cps.gov.uk