Crown Court Data Assurance

Ministry of Justice

December 2024

Contents

Report Status

Closing meeting: 29/11/2024

Draft Report issued: 03/12/2024

Revised Draft Report issued: 05/12/2024

Final Draft issued: 06/12/2024

Final Report issued: 10/12/2024

01 - Executive Summary

Background

The Ministry of Justice (MoJ) is responsible for Courts, Tribunals, Prisons, Probation services and Attendance Centres. HM Courts & Tribunals Service (HMCTS) is an executive agency of MoJ and is responsible for the administration of criminal, civil and family courts in England and Wales, as well as the reserved unified tribunals across the United Kingdom. The Department is seeking an independent review of the methodology and process it uses to produce the Crown Court caseload statistics. This is in response to a discrepancy identified in the Crown Court open caseload data in June 2024 which resulted in the suspension of publication of the Criminal Court Quarterly Statistics. Remedial work has been undertaken internally to address the issue. The Criminal Court Quarterly Statistics highlight the type and volume of cases received and processed through the criminal court system of England and Wales, including statistics on case timeliness.

Objectives of this assessment

- Provide an overall assessment of the confidence that MoJ should have in the Crown Court caseload statistics.

- Review the design of the methodology used by the MoJ for the compilation of Crown Court caseload statistics, including the approach to defining cases and calculating the size of the open caseload.

- Review the operational validation of cases that HM Courts & Tribunals Service (HMCTS) are undertaking, and Quality Assurance (QA) processes in place at both HMCTS and MoJ, to inform confidence levels in the statistics being presented.

Conclusion

- **The MoJ can have a significant level of confidence in the Crown Court caseload statistics, though we have identified minor improvement opportunities.**

We have reviewed the design of the methodology used by the MoJ for the compilation of Crown Court caseload statistics, including the approach to defining cases and calculating the size of the open caseload. We have reviewed the operational validation of cases that HM Courts & Tribunals Service (HMCTS) are undertaking, and Quality Assurance (QA) processes in place at both HMCTS and MoJ, to inform confidence levels in the statistics being presented. We have detailed the findings arising in these areas in Section 02 of this report.

This assessment has been informed by our understanding of the processes and controls in place at the time of this review, and not from validating the caseload data, or the logical integrity of the data pipeline (i.e. code base). We have gained our understanding of the processes and controls through the following activities:

- Interviews with key stakeholders in HM Courts & Tribunals Service (HMCTS) and MoJ analytical teams, and Operations Crime Service Team;

- Reviewing the MoJ logic map detailing data inputs, transformation, to outputs (data architecture map) across the Xhibit, Common Platform and MIS databases, including the open cases (and receipts and disposals);

01 - Executive Summary (cont.)

Conclusion (cont.)

- Walkthroughs of the data pipeline from case management systems to data teams through to publication against documented processes;

- Undertaking an assessment of whether data processing and analytics undertaken by data teams is in accordance with documented processes;

- Walkthroughs and limited sample testing of the Quality Assurance processes and operational validation undertaken by both HMCTS and MoJ;

- Undertaking a desktop review of OneCrown project documentation and review of existing processes to confirm alignment with the OneCrown methodology; and

- The review and consideration of internal remedial work completed to date to rectify historic issues in the Crown Court data.

This review has not considered those areas deemed 'out of scope' as detailed on slide five.

Findings and Recommendations

In Section 02 of this report ('Findings and Recommendations'), we have identified seven 'Low' rated findings. These 'Low' rated findings are considered risks that MoJ should seek to address by the target date but, have no material impact on the confidence MoJ can have in the Crown Court caseload statistics. During the course of our work, we have identified no 'Medium' or 'High' rated findings that would impact on the confidence MoJ can have in the Crown Court caseload statistics.

In Section 03 of this report ('Continuous Improvement Recommendations'), we have identified eight recommendations that MoJ should consider implementing by way of good practice.

01 - Executive Summary (cont.)

Out of scope

The scope of this review has included the processes and controls in place regarding the production of Crown Court caseload statistics. Methodology regarding the data and statistics relating to other criminal courts has not been reviewed.

The assessment of the confidence that the MoJ can have in the Crown Court caseload statistics is limited to the activities undertaken during this review (as noted in Appendix A) and the processes and controls in place at the time of the review, and does not provide an assessment of our confidence in future activities or controls implemented by MoJ over the quality of Crown Court caseload data.

We have not completed detailed testing of the operating effectiveness of the controls in place for the compilation of Crown Court caseload statistics.

We have not completed any validation testing on Crown Court caseload data and therefore have not provided a judgement on whether the caseload data is free from error or on the completeness or accuracy of the data. Likewise, we have not assessed the logical integrity or the 'verification of' the supporting data pipeline and its associated code base.

This review has not included an evaluation of the IT General Controls (ITGCs) across the Crown Court caseload data pipelines. We have therefore not assessed the design or effectiveness of the controls that may support the pipeline in the preparation and publication of the Crown Court caseload statistics.

We have not undertaken any review of the Code supporting the Crown Court caseload data pipelines.

02 – Findings and Recommendations

Dedicated Data Pipeline Testing Team			
Low	**Finding:**	**Risk:**	**Responsible entity:**
2.1	The quality assurance of newly published code to the data pipeline is conducted through peer review. It was confirmed that another member of the Data Science team reviews the developed code to ensure it is fit for purpose and free from error, prior to publishing to the pipeline.	There is a risk that code base review may not be conducted as effectively as it could be. Potentially erroneous code could be pushed through development environments without sufficient, high quality review from individuals who are Suitably Qualified, Experienced Personnel (SQEP).	HMCTS MoJ **Target date:** 30 September 2025
	The approach often used by industry (i.e. software houses) and wider government is that separate teams perform the develop and review roles. The review team is often staffed by individuals who are dedicated testers or model reviewers, who have undergone specific training and are highly experienced in review.	**Recommendation:** HMCTS and the MoJ should treat development and model review responsibilities as distinct and separate roles.	
	We found that there are no dedicated testers working on the pipeline. Instead, testing is undertaken by developers as part of their core remit.	Assign dedicated testers, who could be drawn from other build teams to collaborate in evaluating changes to the pipeline code base, with formal training given on how to review code, assessing it for error and best practice. The dedicated testers should also collaborate with the build team to continually improve the integrity of the data pipeline.	

Enable Data Encryption at rest on the AP			
Low	**Finding:**	**Risk:**	**Responsible entity:**
2.2	Data is not encrypted at rest on the Analytical Platform (AP).	Best practice would be to enable this setting as an additional layer of security and prevention of data misuse.	MoJ **Target date:** 31 March 2025
	Encryption of data at rest ensures that unauthorised users cannot read, mistakenly change, or tamper with the data if they were to gain access.	We confirmed that existing controls (e.g. single sign-on, role-level permission) are in place to mitigate the risk of amending or viewing sensitive data.	
	We confirmed that the AP had been defaulted to not encrypt at rest, though this was a function that could be enabled.	**Recommendation:** The MoJ should ensure that Encryption is enabled at rest on the AP. A review should be conducted of encryption practices across both platforms to verify compliance and identify any gaps.	

02 – Findings and Recommendations (cont.)

OneCrown Methodology Steering Group Governance			
Low 2.3	**Finding:** There is no formal process for obtaining assurance over the implementation of agreed actions following decisions made regarding OneCrown methodology changes by the OneCrown Steering Group. It was confirmed that a process exists to discuss, agree upon, and document decisions made regarding methodology changes at the OneCrown Steering Group. However, no formal process was identified to follow up and verify the implementation of the agreed actions.	**Risk:** There is a risk that agreed-upon actions regarding changes to the OneCrown Methodology are not implemented effectively or within the required timeframe prior to publication of the statistics. This could lead to an inconsistent application of agreed-upon methodologies, resulting in potential errors in the methodology itself and ultimately in the production of Crown Court caseload data and statistics. **Recommendation:** HMCTS and MoJ should establish a formal process for obtaining assurance over the implementation of agreed actions by the OneCrown Steering Group. This process should incorporate steps to verify that the agreed-upon changes have been implemented correctly and in accordance with the agreed-upon timeline. This could be completed through establishing a feedback loop between the Technical Governance forum and the OneCrown Steering Group.	**Responsible entity:** HMCTS MoJ **Target date:** 31 March 2025

Full Coverage of the Data Dictionary for the Data Pipeline			
Low 2.4	**Finding:** We found that data dictionaries are used to support the data pipeline's understanding. However, it was suggested that these are used inconsistently throughout the end to end data pipeline, primarily on the MoJ side, whereby not all fields are complete. This potentially hinders data interpretation and the management of fields, and increases likelihood of selecting incorrect fields for use in the code base.	**Risk:** There is a risk of inaccurate data interpretation and data field management challenges due to inconsistent data definitions and inconsistent data utilisation. **Recommendation:** The MoJ should implement a consistent data dictionary approach across all fields in the data pipeline.	**Responsible entity:** MoJ **Target date:** 30 June 2025

Formalised Model Review Process in Data Pipeline			
Low	**Finding:**	**Risk:**	**Responsible entity:**
2.5	The data pipeline is not subject to a formalised process for implementing time-bound review points, such as annual reverification, in accordance with government best practices. Using strict definitions, the data pipeline might not be regarded as a "model" under defined standards within the likes of the Greenbook, McPherson, or Aquabook. As a result, it wouldn't attract certain handling conditions such as time-bound review points. However, this data pipeline supports key reporting and decision-making processes across MoJ. As such the data pipeline should be regarded as a business critical asset and subject to regular review and verification.	There is a risk that the data pipeline may become outdated, less-accurate, or erroneous over time. Improvements and updates will be made on a continuous basis to the data pipeline. Without a full, regular end-to-end review of the data pipeline, there is a possibility that smaller (localised) changes may not consider the macro impact to the data pipeline. For example, changing data structures or units of measurement that are used in later parts of the pipeline, unknowingly influencing its integrity. Existing code review of these local changes will provide a control to this, but they will often not consider the macro impact to the data pipeline (as it's not often in scope). Hence a regular review will act as an additional mitigating control. **Recommendation:** The data pipeline should be reviewed and verified regularly, with a schedule and scope to be defined by the director level governance forum and containing both MoJ and HMCTS.	HMCTS MoJ **Target date:** 30 September 2025 and then annually

02 – Findings and Recommendations (cont.)

Outstanding Cases Being Erroneously Reported

Low	Finding:	Risk:	Responsible entity:
2.6	From the most recent data snapshot of Technical Exceptions (i.e. cases that do not comply with the expectations of Xhibit), a daily report is generated across all data points in the system. The most recent exception report identified 1,280 cases that required further investigation (out of a total caseload of c.60,000). Of this, 1,249 were actioned and confirmed as corrected and so appear correctly in the caseload data. At the time of this review, there were 31 cases remaining that did not comply with the expectations of the Platform and impacted the criminal caseload count. These cases are closed, however are appearing as open in the caseload data. This is a relatively small number of erroneous reports that have been identified by HMCTS's controls. There's ongoing work to resolve the issues causing these cases to be incorrectly reported.	At the time of this review, HMCTS were continuing to investigate these 31 cases to eradicate the risk of further cases with these characteristics arising. The investigation had not yet been completed. **Recommendation:** HMCTS need to correct the issues leading to the erroneous 31 remaining cases. HMCTS should continue to run data quality reporting to ensure future issues do not arise. Common Platform aim to reduce defects and reach steady state by end of 2025.	HMCTS **Target date:** 31 December 2025

Duplicate Case Detection and Data Management Accuracy

Low	Finding:	Risk:	Responsible entity:
2.7	Cases could, in theory, be entered twice, once in Xhibit and once in Common Platform, without an automated detection system. Despite this, the occurrence of duplicate entries is expected to be rare, requiring someone to enter information in both systems twice, and such errors are unlikely to advance through the system unnoticed. Three controls are in place to mitigate this: 1. Local court MI is reviewed regularly which would spot this duplication (it would only be possible at a single court level); 2. Involved parties to the case would also see duplicate case entries and likely question its duplication; and 3. Reports are generated on cases that have no activity for more than three months, to which these are then investigated to understand reasons why.	There is a risk that caseload data figures could be inflated due to the ability to enter cases twice, once in Xhibit and once in Common platform, and not identified due to the lack of a detective control. **Recommendation:** HMCTS should establish an additional control, at the data pipeline level, to identify instances of duplicate cases across Xhibit and Common Platform. Formal resolution procedures should be established to rectify any duplicates identified through this check.	HMCTS **Target date:** 31 January 2025

03 – Continuous Improvement Recommendations

The following Continuous Improvement recommendations have been identified in the course of this review. Whilst the observations and associated recommendations identify improvements that can be made to the processes or controls in place, they do not directly impact the caseload data figures.

Continuous Improvement Recommendations			
Ref	**Title**	**Observation**	**Recommendation**
3.01	**Addressing Data Validation Issues**	The MoJ data modelling team reviews projected caseload figures against actual outcomes observed. It was suggested they maintain a 'low tolerance' (in the hundreds, i.e. <0.5% of overall caseload) of variance to warrant investigation, however this is not formally defined. The same team receives data feeds weekly, and reviews and investigates these variances more regularly than the quarterly cycle of the publishing of Crown Court caseload statistics, allowing issues to be addressed early. When a significant variance is detected, it is subject to comprehensive investigation by the team who are attempting to establish the root cause of the variance, before the release of any statistical data. This process serves as a control mechanism to scrutinise unexpected spikes or drops in the data, ensuring accuracy and reliability of the published statistics. The existing control can not identify systemic issues (e.g. erroneous code or consistent user input error) that would impact both projected figures and actual outcomes observed.	The tolerance level that triggers an investigation into the Crown Court caseload data by the data modelling team should be formally defined.
3.02	**Automating Data Pipeline to Reduce Errors**	The current procedure for initiating the data pipeline involves a manual process of extracting data from the Common Platform and subsequently re-uploading it to the SDP system. This manual intervention introduces a potential risk for user error, which could impact the accuracy and integrity of the data being processed. This risk is mitigated by the fact that the data pipeline will fail to run if the data is not in the correct format or structure it expects. It is also mitigated by teams who analyse variances (i.e. data modelling team mentioned in Ref 3.01). The data engineering team has acknowledged the risk of manual data transfer and is currently exploring options to automate this step in the process. Automating this procedure could significantly reduce the risk of user error and enhance the efficiency and reliability of the data pipeline.	The progress towards automation should be closely monitored, and resources be allocated to support the implementation of a solution to mitigate the identified risk.

03 – Continuous Improvement Recommendations (cont.)

Continuous Improvement Recommendations			
Ref	Title	Observation	Recommendation
3.03	**Random sampling to become a BAU activity**	As part of validating the data pipeline in response to the issues previously found in the statistics, a one-time quality assurance exercise involving random sampling of 1,000 cases per region was conducted (7,000 cases total out of a caseload total of c.60,000). This exercise aimed to ensure the accuracy and reliability of the data captured at a user entry level. The exercise confirmed a 98% accuracy level in the recorded cases, and the 2% of cases deemed not to be accurate were then edited to align to the correct 'open' or 'closed' status. At the time of this review, discussions were ongoing to determine if a random sampling approach to quality assurance on the Crown Court caseload data could be incorporated into business as usual (BAU) operations. The completion of such quality assurance activities provides assurance over the accuracy of the caseload data captured at a user entry level.	HMCTS should incorporate a random sampling approach to quality assurance on the Crown Court caseload data as part of BAU quality assurance activities.
3.04	**Reconciliation of MoJ / HMCTS branches of the Data Pipeline to a single feed**	The MoJ & HMCTS currently operate two versions of the data pipeline, namely SDP and AP, both of which are used to calculate caseload statistics. Despite reportedly employing the same methodology and mirrored code base, these pipelines are executed on different systems. It was reported that both pipelines yield identical caseload results, and any modifications to one pipeline are reflected in the other, with substantive changes subject to approval by the data decision committee, and then executed in each system. This dual-pipeline approach ensures consistency in caseload calculations across different technological platforms. However, the process of mirroring changes between the two pipelines could introduce delays or discrepancies if not managed appropriately, and is not considered good practice due to having to maintain and run duplicate data pipeline branches.	The existing dual-pipeline approach and resulting duplication of code should be streamlined into a single approach. It should be noted that data science teams already acknowledge this needs to occur, and intend to complete a reconciliation between the two pipelines by March 2025.

03 – Continuous Improvement Recommendations (cont.)

	Continuous Improvement Recommendations		
Ref	**Title**	**Observation**	**Recommendation**
3.05	**Establishing Data Pipeline Governance**	A project-specific governance model has been stood up in response to issues identified in the data pipeline. However, it has yet to be confirmed what this governance model will look like as the project moves forward and new and/or planned controls are implemented into the data pipeline as the project progresses.	The current project implementation governance is a single slide diagram, a more substantive governance model for the project should be developed and implemented to outline clear roles, responsibilities, and processes for ongoing management and oversight of the data pipeline, ensuring its continued effectiveness and alignment with organisational objectives.
3.06	**Pipeline Documentation Under Development**	The infrastructure and procedural documentation for newly changed parts of the data pipeline code base (e.g. instructions, architecture diagrams) were under development at the time of this review. We confirmed it was not mandated that newly onboarded staff were provided with access to existing documentation before working on the pipeline and training arrangements were informal.	All relevant documentation pertaining to the data pipeline should be made available to those being onboarded to work on or develop the pipeline. As part of the newly established onboarding process, documentation and formal training should be given to new users, and their understanding checked before they undertake work on the data pipeline.
3.07	**Formalising the MoJ Model Senior Responsible Officer (MSRO) role**	For HMCTS, there is a designated MSRO for criminal caseload statistics. We were informed that for the pipeline of caseload statistics for MoJ, the MSRO role is not formally defined or well-communicated. We noted through interviews with MoJ stakeholders, that a colleague was suggested as the MSRO for the caseload statistics for MoJ, though there was uncertainty behind this.	Clearly define, document and raise awareness of the roles and responsibilities of the MSRO for criminal caseload statistics on the MoJ side. Ensure regular communication and updates to maintain clarity, accountability, and alignment across teams.
3.08	**User Access Management and Automation**	Role-level access to AP and SDP was reported to be reviewed on a six-monthly basis. AP user access expires after six months, and removes administration-level access upon staff rotation/departure. The SDP has a manual leaver process, with an automated system in development.	Implement a structured, automated process for the regular review and updating of user roles and permissions across AP and SDP to ensure they remain aligned with business requirements and security policies. With this six-monthly review of user permissions ideally occurring three-monthly.

04 - Disclaimer

This Report has been prepared on the basis set out in our Call Off Order Form to the Framework Contract RM6188 Lot 4 (reference: ITT_9037) as set out in the contract (reference: Con_24522) dated 18 November with the Ministry of Justice (the "Client") and its attachments (together the "Agreement"),and should be read in conjunction with the Agreement. We have not conducted an audit and accordingly the scope of our work is different from that of an audit and does not provide the same level of assurance as an audit. This engagement is not an assurance engagement conducted in accordance with any generally accepted assurance standards and consequently no assurance opinion is expressed. We have not verified the reliability or accuracy of any information obtained in the course of our work, other than in the limited circumstances set out in the Agreement. Nothing in this report constitutes a valuation or legal advice.

This Report is for the benefit of the Client only. This Report has not been designed to be of benefit to anyone except the Client. In preparing this Report we have not taken into account the interests, needs or circumstances of anyone apart from the Client, even though we may have been aware that others might read this Report. We have prepared this report for the benefit of the Client alone. This Report is not suitable to be relied on by any party wishing to acquire rights against KPMG LLP (other than the Client) for any purpose or in any context. Any party other than the Client that obtains access to this Report or a copy (under the Freedom of Information Act 2000, the Freedom of Information (Scotland) Act 2002, through the Client's Publication Scheme or otherwise) and chooses to rely on this Report (or any part of it) does so at its own risk. To the fullest extent permitted by law, KPMG LLP does not assume any responsibility or liability in respect of this Report to any party other than the Client.

Appendix A: Approach

We have completed the following activities in the course of our work:

- Interviews with key stakeholders in HM Courts & Tribunals Service (HMCTS) and MoJ analytical teams, and Operations Crime Service Team;

- The development of a logic map of our understanding of data inputs, transformation, to outputs (data architecture map) across the Xhibit, Common Platform and MIS databases, including the open cases (and receipts and disposals);

- Walkthroughs of the data pipeline from case management systems to data teams through to publication against documented processes;

- An assessment of whether data processing and analytics undertaken by data teams is in accordance with documented processes;

- Walkthroughs and limited sample testing of the Quality Assurance processes and operational validation undertaken by both HMCTS and MoJ;

- Desktop review of OneCrown project documentation and review of existing processes to confirm alignment with the OneCrown methodology; and

- The review and consideration of internal remedial work completed to date to rectify historic issues in the Crown Court data.

Appendix B: Documentation reviewed

Review Area	Documentation Reviewed
Quality assurance at source/operational QA	20240910-One Crown data quality and data validation 200240925-OSCG One Crown data validation
Flow to SDP/data pipeline	Data Architecture.png crown_court_full_ERD SDS RMDP SDP to MOJAP Data Pipeline HLD v0.2 RMDP - Curation HLD v1.2 RMDP SDP Data Ingestion V2.1(2)
OneCrown methodology	20240805-OneCrown data PoaP 20240712-OneCrown Court Data PID OneCrown Data Modelling 20240902-One Crown Data rule decision log OneCrown-conceptual ERD crown_court_star_schema
OneCrown governance	20240902-One Crown Data rule decision log 20240726-Data development principles HMCTS TAB ToR v1.6
Processes (e.g. Most serious disposals, bench warrants etc.)	SDP Onboarding SDP RBAC HMCTS SDP Data Access Processs - Version 3.0
Flow into HMCTS Strategic data platform	Data Architecture.png crown_court_full_ERD

Appendix C: Findings - Ratings definitions

We have set out below the priority ratings used to assess each individual finding.

Priority	Description
Red – High	A significant weakness in the processes and controls in place regarding the production of Crown Court caseload statistics which severely limits the confidence MoJ can have in the statistics. Any findings and management actions in this category would require immediate attention before the publication of the Crown Court caseload statistics.
Amber – Medium	A potentially significant or medium level weakness in the processes and controls in place regarding the production of Crown Court caseload statistics which limits the confidence MoJ can have in the statistics. If left unaddressed, findings and management actions in this category would put MoJ at risk of not having the required confidence in the statistics to publish the Crown Court caseload statistics.
Green – Low	Management actions which could improve the efficiency and / or effectiveness of the processes and controls in place regarding the production of Crown Court caseload statistics. These are generally issues that would not materially impact the confidence MoJ can have in the Crown Court caseload statistics prior to their publication.

kpmg.com/uk