

Appariement des données administratives des services de sécurité intérieure et de la justice

Le cas des violences conjugales

Malik Koubi, Nicolas Razafindratsima, SSER*
Antonin Briand, Nicolas Claude, SSMSI**

Synthèse

La chaîne pénale se partage entre le ministère de l'Intérieur (MI) et celui de la Justice (MJ). Pour des raisons historiques qui tiennent à leurs missions respectives, ces deux administrations ont développé chacune leur propre système d'information qui sont aujourd'hui peu interoperables. La reconstitution de l'intégralité du traitement pénal des affaires présente pourtant un enjeu pour plus d'efficacité, de cohérence et de transparence dans la production statistique. Son objectif est aujourd'hui d'obtenir un identifiant unique pour les procédures/affaires associé à la procédure pénale numérique (PPN), dispositif lancé en 2018 de dématérialisation des pièces et de transmission automatique des procédures du MI au MJ. Néanmoins, la PPN n'est pas encore pleinement opérationnelle à ce jour.

Le projet d'appariement de données des deux services statistiques ministériels de la Sécurité intérieure (SSMSI) et de la Justice (SSER, Service de la statistique, des études et de la recherche) répond à une demande exprimée par de nombreux partenaires ces dernières années, en particulier les acteurs de la statistique publique. Elle figure ainsi dans les avis moyens terme du Conseil national de l'information statistique (Cnis) (moyen terme 2019-2023 puis 2024-2028). Elle a également été exprimée par l'Autorité de la statistique publique dans un délibéré datant de 2022. Les deux services statistiques ministériels collaborent activement depuis cette date pour expertiser et préparer leurs données administratives afin qu'elles se prêtent à l'opération d'appariement.

Ce dossier, réalisé par le SSER et le SSMSI, présente le premier travail de rapprochement statistique portant sur l'ensemble de la chaîne pénale, du constat de l'infraction par les services de sécurité intérieure jusqu'à la décision de justice. Il porte sur le champ infractionnel des violences conjugales. Les sources administratives issues des applicatifs des deux ministères sont mobilisées : les données issues des Logiciels de Rédaction de Procédure de la Police Nationale et de la Gendarmerie Nationale (LRPPN et LRPGN) d'une part et les données issues du logiciel de la Chaîne

Applicative Supportant le Système d'Information Orienté Procédure pénale Et Enfants (CASSIOPEE) d'autre part.

Le rapprochement des procédures et des affaires se base sur le numéro du procès-verbal, inscrit sous forme d'un triplet de variables administratives (Année de la création de la procédure / Code de l'unité rédactrice de la procédure/ Numéro de la procédure) qui est transmis aux parquets. Toutefois, il n'est pas systématiquement présent dans l'applicatif Cassiopée (transmission imparfaite ou retardée, non saisie, erreur de saisie, etc.). Avec le triplet uniquement, le taux d'appariement obtenu est d'environ 75 %. Un rapprochement additionnel est réalisé à partir du code statistique non signifiant (CSNS) des personnes mises en cause dans les deux systèmes d'information. Ce second niveau d'appariement améliore le taux d'appariement de 3 à 5 points selon les années.

Une partie des procédures de l'Intérieur sont liées car elles se rapportent au même dossier. Le regroupement de ces procédures en paquets et le changement d'unité statistique qui en résulte contribue à améliorer le taux d'appariement de 2 à 3 points. Finalement, le taux d'appariement optimisé des procédures aux affaires s'élève à 81 %.

Le défaut de représentativité découlant de ces quelque 20 % de procédures non-appariées est enfin évalué à l'aide d'une analyse factorielle (analyse en composantes multiples). Le facteur explicatif principal du non-appariement est l'absence de mis en cause identifié. Viennent ensuite des effets liés à des établissements particuliers (comme la préfecture de Police de Paris) et des effets géographiques.

Dans ce dossier, une première partie décrit les sources de chacun des deux services statistiques. Est ensuite présentée l'expertise menée sur les données de la justice ayant servi à réaliser le rapprochement des sources. Dans une troisième partie, la méthode de constitution des paquets de procédures dans les bases du SSMSI est présentée, suivie des résultats de l'appariement mis en œuvre sur le champ des violences conjugales. Enfin, une dernière partie dresse les premiers constats relevés sur les procédures non appariées.

* Nous remercions l'équipe Cassiopée du BSPP pour leur appui : Pierre Jung, François Malet, Thibaud Meynier, Wali Rostam et Léonie Veuille.

** Expertise des bases de données du ministère de l'Intérieur par Laurianne Salembier, SSMSI.

Sommaire

Introduction	4
Présentation des données des deux services statistiques	5
Données du service statistique du ministère de l'Intérieur (SSMSI)	5
Données du service statistique du ministère de la Justice (SSER)	6
Expertise des données Cassiopée en vue de l'appariement avec les données du ministère de l'Intérieur	7
Le triplet, un lien central entre les données des deux ministères	7
L'expertise de deux tables Cassiopée de données « primaires »	7
Le rapprochement des deux tables primaires donne une clé d'appariement entre les deux systèmes d'information	9
L'enrichissement par le Code statistique non signifiant (CSNS)	9
Constitution des paquets de procédures dans les bases du SSMSI	10
Création des couples de procédures avec la procédure d'origine	10
Création des couples de procédures avec l'identité des victimes et des mis en cause	11
Regroupement des méthodes et création des paquets de procédures	11
Appariement sur caractéristiques communes sur le champ des violences conjugales	12
Appariement sur triplet et validation des appariements	13
Appariement sur CSNS et validation des appariements	14
Regroupement des deux méthodes d'appariement	15
Apport des paquets de procédures pour l'appariement	15
Analyse des relations multiples entre les procédures et les affaires	16
Caractérisation des procédures appariées et non-appariées sur le champ des violences conjugales	17
Identification des variables par analyse des correspondances multiples (ACM)	17
Analyse de la présence d'un mis en cause dans la procédure sur l'appariement	18
Analyse de l'impact de l'enregistrement de la procédure par la préfecture de Police de Paris ou par la DTPN sur l'appariement	19
Conclusion	19
Pour en savoir plus	20

Liste des figures

Figure 1.	Le code statistique non significatif (CSNS)	6
Figure 2.	Nombre d'affaires par procédure.....	8
Figure 3.	Nombre de procédures par affaire	8
Figure 4.	Proportion de triplets valides selon l'année d'enregistrement et l'origine de l'affaire (en %)	9
Figure 5.	Répartition des mis en cause de Cassiopée selon qualité du CSNS.....	9
Figure 6.	Répartition des couples de procédures de la police nationale selon la concordance de la commune de commission et de la date du fait (en %)	10
Figure 7.	Répartition des couples de procédures de la gendarmerie nationale selon la concordance de la commune de commission et de la date du fait (en %)	10
Figure 8.	Théorie des graphes pour regrouper les procédures.....	11
Figure 9.	Distribution de la taille des paquets de procédure de la police nationale.....	12
Figure 10.	Distribution de la taille des paquets de procédure de la gendarmerie nationale.....	12
Figure 11.	Schéma représentatif de l'appariement de la procédure à la réponse pénale.....	12
Figure 12.	Définition du champ des violences conjugales dans les données issues des FSI.....	12
Figure 13.	Représentation de l'appariement entre les procédures et les affaires	13
Figure 14.	Constitution de la clé d'appariement selon l'origine de la procédure (FSI).....	13
Figure 15.	La qualité du triplet dans Cassiopée pour le champ des violences conjugales	13
Figure 16.	Taux d'appariement via le triplet des procédures de violences conjugales avec au moins une affaire Cassiopée entre 2018 et 2023	14
Figure 17.	Exemple d'appariement via le CSNS	14
Figure 18.	Taux d'appariement via le CSNS des mis en cause des procédures de violences conjugales avec au moins une affaire Cassiopée entre 2018 et 2023	15
Figure 19.	Taux d'appariement via le triplet et le CSNS des mis en cause des procédures de violences conjugales avec au moins une affaire Cassiopée entre 2018 et 2023 (en %)	15
Figure 20.	Comparaison des taux d'appariement avec au moins une affaire Cassiopée des procédures de violences conjugales et de celui des paquets de procédures conjugales entre 2018 et 2023 (en %)	16
Figure 21.	Relations multiples possibles	16
Figure 22.	Exemple d'une relation de deux procédures de deux paquets différents appariées à une même affaire.....	16
Figure 23.	Exemple de cas de deux affaires appariées à la même procédure	17
Figure 24.	Analyse des correspondances multiples (ACM)	17
Figure 25.	Statut d'appariement des procédures 2022 de police nationale selon la présence d'un mis en cause (en %).....	18
Figure 26.	Statut d'appariement des procédures 2022 de gendarmerie nationale selon la présence d'un mis en cause (en %)	18
Figure 27.	Statut d'appariement des procédures 2022 de police nationale selon l'enregistrement par la préfecture de Police de Paris (en %)	19
Figure 28.	Statut d'appariement des procédures 2022 de police nationale selon l'enregistrement à la DTPN (en %)	19

La procédure pénale désigne l'ensemble des règles et des étapes qui encadrent la recherche, la poursuite et le jugement des auteurs d'infractions. Elle fait intervenir deux administrations différentes que sont le ministère de l'Intérieur via les forces de sécurité intérieure (FSI, police nationale et gendarmerie nationale) et le ministère de la Justice, ainsi que leurs services statistiques, le SSMSI et le SSER. La reconstitution du déroulement de la chaîne pénale, du constat de l'infraction ou du dépôt d'une plainte jusqu'à la décision pénale, répond à une demande sociale et politique importante, exprimée notamment dans l'avis à moyen-terme 2019-2023 du CNIS et rééditée dans l'avis 2024-2028. L'Autorité de Statistique Publique recommande pour sa part dans un délibéré du 14 février 2022 de mettre en place « *un suivi longitudinal conjoint des procédures pénales, retraçant, indépendamment des variations de leur qualification juridique, leur évolution du début à la fin de leur vie judiciaire.* »

La procédure pénale numérique (PPN), projet interministériel, a été lancée en 2018 par le ministère de l'Intérieur et le ministère de la Justice dans le but de réaliser un traitement intégré des affaires pénales. Il s'agit de dématérialiser le processus de traitement, depuis l'acte d'enquête initial (enregistrement de la plainte, du constat ou du signalement) jusqu'à l'archivage de la décision de justice, en s'appuyant sur la mise en place d'un identifiant unique entre les procédures et les affaires. La PPN permet aujourd'hui d'améliorer la qualité des données mais, à ce stade, sa montée en charge n'est pas achevée. Fin 2024, trois quarts des affaires classées sans suite étaient transmises via la PPN contre seulement un tiers des affaires non classées sans suite.

Le travail de rapprochement statistique des deux parties de la chaîne pénale constitue donc un investissement nécessaire en attendant la mise en place de l'identifiant unique dans le cadre de la PPN. Il mobilise les sources administratives issues des applicatifs des deux ministères : les données issues des Logiciels de Rédaction de Procédure de la Police Nationale et de la Gendarmerie Nationale (LRPPN et LRPGN) d'une part¹, et les données issues du logiciel de la Chaîne Applicative Supportant le Système d'Information Orienté Procédure pénale Et Enfants (CASSIOPEE) d'autre part. Ces sources sont hétérogènes car chaque ministère possède, en lien avec ses missions et ses contraintes, sa propre façon d'enregistrer et de structurer l'information.

Cette étude est le premier travail de reconstitution au niveau le plus fin des procédures pénales dans leur intégralité, du constat de l'infraction ou du dépôt d'une plainte à la réponse pénale, sur le champ infractionnel des violences conjugales. Ce champ a été choisi pour cette première étude car il répond à une attente sociale forte et a fait l'objet de politiques très volontaristes ces dernières années. En outre, les affaires s'y rapportant ont l'avantage de porter sur des faits engageant un nombre limité de personnes, et qui sont clairement désignées (dans la majorité des cas). Par exemple, dans la majorité de ces affaires, une personne mise en cause est bien identifiée, i.e. une personne à l'encontre de laquelle il existe des

raisons plausibles de soupçonner qu'elle ait commis ou ait tenté de commettre une infraction.

Cette étude s'inscrit dans la lignée d'un certain nombre de travaux entrepris par les services statistiques, d'abord à un niveau agrégé. Il s'agissait alors de réconcilier les niveaux et les évolutions des infractions liées aux stupéfiants (Groupe de travail SSMSI SDSE, 2016), ou encore des violences conjugales (Brunin, Guedj, & Le Rhun, 2019) vus par les deux systèmes statistiques. L'analyse menée à ce moment-là a pu préciser la façon dont les différences de définition d'unités de compte, de périmètres, de nomenclatures, les différences de temporalité et les requalifications des faits conduisent à de sérieuses divergences entre les données intérieur et les données justice. Des corrections ont été identifiées et peuvent rapprocher les chiffres, parfois même de façon très satisfaisante comme pour les violences conjugales, mais ces corrections ne suffisent pas à elles seules à donner une vision unifiée des évolutions de la délinquance.

Les premiers travaux de rapprochement au niveau fin des données intérieur et des données justice ont été menés sur le champ des homicides et des tentatives d'homicide (Makdessi & Carrasco, 2022), mais n'ont pas pu aboutir complètement : la proportion de faits appariés entre les deux bases de données apparaissait relativement faible (55 % sur l'année 2020). Les deux services ont considéré que ce champ infractionnel était trop spécifique du fait de sa grande porosité avec les coups et blessures volontaires et de la multiplicité des procédures se rapportant au même dossier côté intérieur, de nombreux services intervenant dans l'enquête (Salember, 2024a).

L'unité statistique des données issues des systèmes d'information des FSI est la **procédure**. Concept administratif, elle rassemble l'ensemble des actes d'enquête dressés par un service au cours d'une saisine donnée et est transmise à la justice comme un ensemble de pièces à conviction. Une saisine d'un service de sécurité intérieure peut être de différentes natures : elle peut être à l'initiative d'un service des FSI, en cas de flagrant délit ou d'un dépôt de plainte par exemple, ou faire suite à une instruction du parquet.

L'unité de base des données issues des systèmes d'information de la justice est l'**affaire pénale**. L'affaire pénale comporte l'ensemble des pièces transmises par les services de sécurité ainsi que les autres éléments concernant son traitement judiciaire. L'affaire pénale devient une unité statistique lorsqu'elle est enregistrée dans Cassiopée. Chaque affaire est alors réduite à un identifiant unique généré automatiquement.

Les procédures et les affaires sont des objets statistiques complexes et composites, parfois difficiles à manipuler car ils comportent plusieurs dimensions. Outre la dimension temporelle et les éléments juridiques et d'enquête, elles impliquent des personnes en nombre variable (victimes, témoins, parties civiles et mis en cause), ayant elles-mêmes leurs caractéristiques propres et leurs antécédents. Les procédures côté FSI, tout comme les affaires côté justice, correspondent en toute généralité à des séquences d'investigations plus ou moins étalées dans le temps. Au fil

¹ Côté ministère de l'Intérieur, sont également mobilisées les données issues des procès-verbaux électroniques et les données saisies sur la plateforme de plaintes en ligne Thésée pour les e-escroqueries.

des investigations, des informations nouvelles émergent et peuvent donner lieu à des ramifications. Ainsi, en cas de dessaisissement d'un service de police ou d'une unité de gendarmerie au profit d'un autre, une nouvelle procédure est ouverte. De même, une nouvelle procédure peut être ouverte si le parquet ou un juge d'instruction demande aux FSI d'accomplir des investigations supplémentaires. L'unité pertinente pour ces procédures successives, correspondant aux différentes saisines, est alors plutôt le « paquet » de procédures, toutes liées à un même dossier. Une procédure peut également se scinder en plusieurs affaires judiciaires, par exemple il arrive qu'une affaire soit enregistrée avec un nouvel identifiant lorsqu'il y a dessaisissement. Certaines affaires peuvent aussi parfois être regroupées si un lien est établi entre elles à un moment au cours du traitement judiciaire. C'est notamment le cas dans les affaires liées à des réseaux de crime organisé dans lesquelles le lien entre les affaires est établi *ex-post*. Enfin au cours de l'enquête, des modifications peuvent intervenir : des personnes mises cause peuvent apparaître, d'autres être mises hors de cause, etc.

La phase de transmission des procédures des FSI vers les parquets est également source de complexité. Elle peut être réalisée sous format électronique ou papier, auquel cas les informations d'identification sont saisies manuellement, bien que ce dernier cas tende à disparaître avec la PPN. Les délais de transmission peuvent également être sources de divergence. C'est en effet à l'issue des investigations des FSI qu'une transmission de la procédure correspondante à la justice a lieu, même si le procureur de la République est tenu informé « *par tous moyens* » dès le début de la procédure.

La stratégie de rapprochement des procédures et des

affaires choisie ici tire parti des différents niveaux d'information présents dans les procédures et les affaires. Chaque procédure des FSI est repérée par un triplet de variables (Année de la création de la procédure / Code de l'unité rédactrice de la procédure/ Numéro de la procédure) qui est en principe transmis à la justice, mais en réalité de manière imparfaite. Il n'est pas systématiquement présent dans Cassiopée (transmission imparfaite ou retardée, non saisie, erreur de saisie, etc.). C'est pourquoi la dimension nominative est également sollicitée sous la forme des traits d'identité des personnes mises en cause par l'intermédiaire de leur code statistique non signifiant (CSNS). Cette stratégie est fructueuse pour réduire les « faux négatifs » mais génère en contrepartie des « faux positifs ». Ils sont réduits à leur tour par une validation à l'aide de variables auxiliaires, la date et le lieu de commission des faits, concepts présents - mais avec des variations - dans les systèmes d'information des deux ministères.

La première partie présente les données provenant des systèmes d'information des ministères de l'Intérieur et de la Justice. La deuxième partie expose le travail de repérage des tables et les variables pertinentes pour l'appariement, côté justice. La troisième partie décrit le travail effectué sur les procédures côté FSI pour détecter celles qui sont liées entre elles et construire des « paquets » de procédures correspondant à un même dossier. La quatrième partie détaille la stratégie de l'appariement entre les données des deux ministères, les traitements statistiques et les contrôles effectués, et présente un bilan de l'appariement. La dernière partie analyse les quelque 20 % de procédures côté FSI qui n'ont pas été appariées avec une affaire côté justice, et présente les caractéristiques principales de ce non-appariement.

Présentation des données des deux services statistiques

Données du service statistique du ministère de l'Intérieur (SSMSI)

Le SSMSI mobilise les données administratives issues des applicatifs du ministère de l'Intérieur, notamment les données issues des logiciels de rédaction de procédures de la police et de la gendarmerie nationales correspondant aux saisies en continu des policiers et des gendarmes². Chaque procédure enregistrée rassemble l'ensemble des actes d'enquête dressés dans le cadre d'une saisine. Cette saisine, par laquelle la justice recourt à un service de police ou à une unité de gendarmerie pour rassembler des preuves dans le cadre d'une enquête judiciaire, peut être à l'initiative du service ou de l'unité, en cas de flagrant délit ou d'un dépôt de plainte par exemple, ou faire suite à une instruction du parquet. En cas de poursuite d'enquête dans le cadre d'une commission rogatoire, cadre d'enquête par lequel le juge d'instruction peut demander à un officier de police judiciaire d'accomplir à sa place un ou plusieurs actes d'instruction, une nouvelle procédure correspondant à cette nouvelle saisine sera ouverte par le policier ou le gendarme. Une fois clôturées, les procédures sont ensuite transmises à l'autorité judiciaire pour poursuite de la chaîne pénale.

Les procédures rédigées par les forces de sécurité intérieure contiennent des informations sur les infractions enregistrées par la police et la gendarmerie nationales, ainsi que sur les éventuelles victimes et mis en cause de ces infractions. Leurs contenus permettent notamment de caractériser les infractions (date et lieu de commission des faits, nature de l'infraction, mode opératoire, etc.), et les personnes associées comme victimes ou comme mis en cause (nom, prénom, date de naissance, sexe, nationalité, etc.). Lorsque les procédures proviennent des données enregistrées par la police nationale par exemple, le cadre juridique de l'enquête est également disponible, tout comme la nature de la saisine ayant donné lieu à la procédure.

Le SSMSI utilise ces données pour constituer des bases statistiques afin d'élaborer des statistiques et des analyses sur la délinquance enregistrée. Trois bases statistiques existent actuellement : la base statistique des victimes, la base statistique des mis en cause et celle des infractions (Salembier, 2024b). Elles couvrent l'ensemble des infractions constatées ou élucidées par la police et la gendarmerie nationales, commises sur le territoire de la république française (France y compris Collectivités

² Le SSMSI exploite également les données issues des procès-verbaux électroniques (y compris les amendes forfaitaires délictuelles) et des plaintes pour e-escroquerie enregistrées en ligne via la plateforme THESEE.

d'Outre-mer). Plus précisément, il s'agit des infractions portées pour la première fois à la connaissance des services de police et des unités de gendarmerie et des infractions élucidées pour la première fois (infractions imputées à un mis en cause entendu).

Dans le cadre de l'appariement avec les données du ministère de la Justice, le SSMSI utilise les millésimes 2018 à 2023 de ces bases statistiques. Il utilise également deux autres types de données :

- des extractions exhaustives contenant l'ensemble des procédures enregistrées par la police et la gendarmerie nationales entre 2018 et 2023, et non uniquement celles entrant dans le champ de ses bases statistiques,
- le CSNS des victimes et des mis en cause apparaissant dans les procédures enregistrées par la police et la gendarmerie nationales entre 2018 et 2023.

En effet, les bases statistiques ne couvrent pas l'ensemble des procédures enregistrées par les forces de sécurité intérieure : certaines procédures enregistrées ne contiennent pas d'infractions constatées pour la première fois, ni d'infractions élucidées pour la première fois. Certains dossiers peuvent en effet faire l'objet de plusieurs saisines pour poursuite d'enquête, par le même service ou par des services différents, et à chaque nouvelle saisine, une nouvelle procédure est ouverte. Or il est a priori possible que ces procédures soient celles qui sont enregistrées dans Cassiopée. C'est pourquoi des extractions exhaustives des données enregistrées dans les logiciels de rédaction des procédures (LRPPN et LRPGN) sont utilisées, contenant l'ensemble des procédures enregistrées (avec ou sans victime enregistrée, avec ou sans mis en cause enregistré, avec ou sans fait constaté pour la première fois). Que ce soit dans les bases statistiques ou dans les extractions exhaustives, les procédures sont caractérisées par un identifiant unique, également présent dans les données sur les affaires issues de Cassiopée, qui constitue la clé d'appariement principale entre les données issues du système d'information du ministère de l'Intérieur et celles issues du système d'information du ministère de la Justice.

Le CSNS des mis en cause et des victimes (**Figure 1**) est utilisé à la fois dans le cadre de la constitution des paquets de procédures et dans le cadre de l'appariement. Un CSNS de qualité 1 à 4 est trouvé pour environ 96 % des victimes apparaissant dans les procédures enregistrées par la police et la gendarmerie nationales entre 2019³ et 2023. Un CSNS de qualité 1 à 4 est trouvé pour environ 88 % des mis en cause apparaissant dans les procédures enregistrées par la police nationale entre 2019 et 2023 et pour 96 % des mis en cause apparaissant dans les procédures enregistrées par la gendarmerie nationale entre 2019 et 2023.

Données du service statistique du ministère de la Justice (SSER)

Cassiopée (Chaîne Applicative Supportant le Système d'Information Orienté Procédure Pénale Et Enfants) est le logiciel utilisé en juridiction pour le traitement des affaires pénales. Le SSER extrait une partie des données « primaires » de l'appliquatif afin de produire des tables

• Figure 1. Le code statistique non significatif (CSNS)

Le code statistique non significatif (CSNS) est un identifiant individuel créé à partir du numéro de sécurité sociale (NIR), haché et chiffré de manière à ne contenir aucune information personnelle signifiante (Bénichou, Espinasse, & Gilles, 2023). L'Insee met à disposition ce service aux services statistiques ministériels demandeurs, dans un cadre juridique strict, leur permettant de récupérer le CSNS des individus de leurs données et de réaliser des appariements sur la base de celui-ci, tout en respectant les règles de confidentialité des données individuelles.

L'obtention du CSNS à partir des traits d'identité des individus (nom, prénom, date et lieu de naissance) se fait en deux étapes :

- Recherche de l'individu dans le répertoire RNIPP (répertoire national d'identification des personnes physiques) à partir de ses traits d'identité.
- Calcul du CSNS à partir du NIR de l'individu trouvé.
- Pour chaque individu, un indicateur de qualité renseigne sur la fiabilité de son identification dans le RNIPP, dont les modalités s'échelonnent de 1 (parfaitement fiable) à 7 (non fiable). Dans le cadre de ces travaux, seuls les CSNS des individus ayant un indicateur qualité compris entre 1 (parfaitement fiable, <1 % de faux positifs) et 4 (assez fiable, <10 % de faux positifs) sont conservés.

statistiques pour le pilotage et les études.

Le logiciel Cassiopée est déployé dans l'ensemble des parquets, tribunaux correctionnels et juridictions pour mineurs, à partir de 2012-2013. Destiné à l'ensemble des services pénaux, du parquet à l'exécution des peines, l'appliquatif Cassiopée vise à traiter les procédures de contravention de 5^e classe, les délits et les crimes (hors phase de jugement) ainsi que les contraventions connexes (commises dans le même temps). Du point de vue de l'information recueillie, les tables statistiques issues de Cassiopée permettent de suivre exhaustivement toutes les affaires pénales dont est saisie la Justice, de leur arrivée au parquet et leur enregistrement⁴ à sa fin, que l'affaire se termine par un classement sans suite, une alternative aux poursuites ou par la décision d'une juridiction de jugement.

À ce jour, Cassiopée permet d'analyser le traitement par le parquet de toutes les infractions reprochées à des personnes physiques (majeures et mineures) ou à des personnes morales enregistrées. Ce système de gestion permet également de connaître la réponse judiciaire pour les décisions en première instance (classement, alternative aux poursuites, renvoi devant une autre juridiction de jugement avec ou sans instruction préalable, jugement) par les tribunaux correctionnels et le mode de saisine du tribunal correctionnel (comparution immédiate,

³ Les CSNS ne sont pas de bonne qualité en 2018 pour la police nationale, la durée de conservation des données nominatives étant de 5 ans et les CSNS ont été produits en 2024.

⁴ Avec le déploiement de la PPN depuis 2021, certaines affaires classées sont automatiquement enregistrées dans Cassiopée, alors que toutes ne l'étaient pas avant cette date. Ces affaires sont classées avant enregistrement au parquet par directive du procureur directement auprès de la police ou de la gendarmerie pour un motif d'absence d'infraction, d'infraction insuffisamment caractérisée ou d'auteur inconnu.

comparution sur reconnaissance préalable de culpabilité, audience correctionnelle sur convocation par officier de police judiciaire). Les décisions des cours d'appel et des cours d'assises ne sont pas actuellement intégrées dans l'appli.

La mise à jour des tables statistiques est réalisée grâce à une remontée d'information effectuée tous les mois, des juridictions vers le SSER. Les données de l'appli Cassiopée sont extraites puis complétées par des informations issues des tables de nomenclatures du Système de Référence Justice (SRJ)⁵, des nomenclatures propres au Système d'Information Décisionnel pénal (SID) et des tables de paramétrage du SID. Ensuite, les données brutes sont soumises à de nombreux traitements spécifiés par le SSER. Elles sont corrigées, transformées et historisées. La table de « faits » est créée : elle donne des

informations détaillées sur l'affaire, les événements, et les personnes mises en cause. Parallèlement, l'enrichissement par le CSNS de la table des mis en cause dans Cassiopée a pu être réalisé en 2024 (voir *infra*). Les données de Cassiopée couvrent l'ensemble de la France (hors Nouméa et Papeete).

La principale unité statistique utilisée dans Cassiopée est l'affaire. Concept juridique, l'affaire correspond dans le logiciel Cassiopée à un dossier contenant le déroulement intégral d'une procédure judiciaire, depuis la saisine du parquet, jusqu'à une décision en première instance, hors crimes. En effet, s'agissant des crimes, l'enregistrement dans le logiciel Cassiopée s'arrête à la décision rendue par le juge d'instruction (ordonnance de non-lieu ou ordonnance de mise en accusation devant la cour d'assises ou la cour criminelle départementale).

Expertise des données Cassiopée en vue de l'appariement avec les données du ministère de l'Intérieur

Pour rapprocher les données des ministères de l'Intérieur et de la Justice, un retour vers les données primaires Cassiopée était nécessaire afin d'y rechercher les données relatives aux procédures transmises par les services de sécurité intérieure vers les parquets.

Si l'information centrale recherchée est la procédure, d'autres informations disponibles permettant de la caractériser ont également été repérées. Des travaux exploratoires et méthodologiques ont été menés en 2024-2025 sur différentes extractions de données réalisées par la direction du numérique du ministère de la Justice. Ces différentes extractions étaient indispensables pour bien appréhender les nouvelles informations issues de Cassiopée et pour déterminer le champ le plus exhaustif possible à retenir pour les rapprochements avec les données du SSMSI.

Le triplet, un lien central entre les données des deux ministères

Les premiers travaux conduits en 2021 et 2022 (Makdessi & Carrasco, 2022) ont permis de repérer un triplet de variables qui caractérisent le procès-verbal établi par les services de sécurité intérieure. Il s'agit de :

- l'année de création de la procédure,
- le code de l'unité rédactrice,
- le numéro de la procédure.

Ces trois champs sont communs à l'appli Cassiopée et aux bases de données des logiciels de rédaction des procédures de la police nationale (LRPPN) et de la gendarmerie nationale (LRPGN) utilisées par le SSMSI. Ce triplet est transmis par les FSI à la Justice lors de la saisine du parquet. Il permet d'identifier une procédure côté FSI, chaque procédure correspondant à une saisine (d'initiative ou sur instruction parquet, notamment). Il est présent dans les données primaires de Cassiopée, mais n'entre pas dans la chaîne de production statistique actuelle du SSER.

L'expertise de deux tables Cassiopée de données « primaires »

La phase de transmission des éléments d'enquête entre le ministère de l'Intérieur et le ministère de la Justice est cruciale pour le travail de rapprochement des données des deux ministères. C'est pourquoi le SSER a mobilisé et expertisé deux tables supplémentaires, dites « primaires », situées en amont de la chaîne de traitement statistique habituelle de Cassiopée. Ces tables primaires contiennent en effet plusieurs informations en provenance du ministère de l'Intérieur, dont le triplet précédent, et sont donc essentielles pour l'appariement. Deux tables primaires ont été repérées comme pertinentes :

- La table ECH_REFERENTIEL_PV : enregistre l'ensemble des saisines (procédures) envoyées par les services de police et de gendarmerie,
- la table AFF_LIEN_AFFAIRE_PV : comporte les liens entre les saisines (procédures) envoyées par les services de police et de gendarmerie et les affaires enregistrées dans Cassiopée.

Elles permettent *in fine* de reconstituer un lien entre le triplet (identifiant une procédure côté forces de sécurité intérieure) et l'affaire judiciaire côté Justice.

La table ECH_REFERENTIEL_PV, un référentiel des saisines reçues par la Justice

La table ECH_REFERENTIEL_PV enregistre toutes les saisines reçues (i.e. les procédures des forces de sécurité intérieure) par les parquets. Chaque ligne de la table est identifiée par une variable [reference_pv_id], qui est un numéro séquentiel. La table contient également le triplet servant aux appariements.

Aperçu de la table primaire Cassiopée ECH_REFERENTIEL_PV

REFERENCE_PV_ID	ANNEE_PROCEDURE	CODE_UNITE_REDACTRIC	NUMERO_PROCEDURE
6606XX	2015	00028XXX	00067
6608XX	2015	00008XXX	00028

⁵ La nomenclature Système de Référence Justice (SRJ) rassemble l'ensemble des tables et nomenclatures de référence du ministère de la justice afin d'avoir un référentiel pénal commun aux applications du ministère et des ministères partenaires.

La table ECH_REFERENTIEL_PV contient 30 650 000 lignes (dernier export en août 2025). Une comparaison avec des versions antérieures de ce fichier (août 2024 et février 2025) montre qu’aucune « purge » n’est effectuée dans cette table. Le stock est régulièrement enrichi des nouveaux procès-verbaux, sans perte d’information sur les années précédentes.

716 000 lignes du fichier (2,3 %) apparaissent d’emblée comme inutilisables pour les appariements, car le triplet est vide ou comporte des valeurs aberrantes (année située en dehors de 2012-2025, ou encore valeurs nulles pour le code service ou le numéro de procédure). En dehors de ces lignes inutilisables, chaque numéro séquentiel n’apparaît qu’une seule fois.

Quelques triplets apparaissent en doublons, mais en si petit nombre (151) que le phénomène semble une anomalie résiduelle. Pour cette raison, un seul triplet a été conservé pour un numéro séquentiel donné, de sorte que la table ECH_REFERENTIEL_PV, une fois traitée de nouveau, établit une correspondance unique entre numéros séquentiels et triplets. La table ECH_REFERENTIEL_PV contient par ailleurs l’ensemble des numéros séquentiels apparaissant dans la deuxième table, AFF_LIEN_AFFAIRE_PV, et peut être considérée à ce titre comme l’identifiant des saisines traitées par les forces de sécurité et reçues par la Justice.

La table AFF_LIEN_AFFAIRE_PV établit une correspondance entre les procédures du ministère de l’intérieur et les affaires du ministère de la Justice

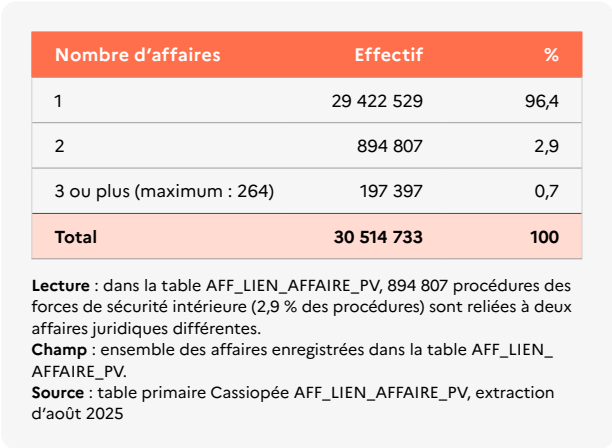
La table primaire AFF_LIEN_AFFAIRE_PV fait le lien entre les saisines des forces de sécurité intérieure (i.e. les procédures) et les affaires au sens juridique. Rappelons que le numéro d’affaire qui se trouve dans cette table est l’identifiant d’affaire présent dans tout le système d’information Cassiopée et permet ainsi de faire le lien avec l’ensemble des informations concernant le traitement judiciaire de l’affaire. La table n’est pas univoque car une affaire peut correspondre à plusieurs procédures côté forces de sécurité intérieure, par exemple celle correspondant à la saisine d’initiative, et celle correspondant à une poursuite d’enquête (commission rogatoire). De même, une procédure côté forces de sécurité intérieure peut alimenter plusieurs affaires : par exemple, si un mineur et un majeur sont mis en cause lors d’une même infraction, le procureur peut ouvrir une affaire distincte pour le mineur et une autre pour le majeur (cela concerne peu de cas).

La table AFF_LIEN_AFFAIRE_PV comprend quatre variables utiles : [Affaire_id] = identifiant de l’affaire juridique, [reference_pv_id] = numéro séquentiel identifiant la procédure côté forces de sécurité intérieure, [date_creation_lien] = date de création du lien entre l’affaire et la procédure, [idj_du_pv] = identifiant justice de l’affaire. Cette dernière variable est remplie pour 5 159 000 observations, soit 16 % des lignes du fichier.

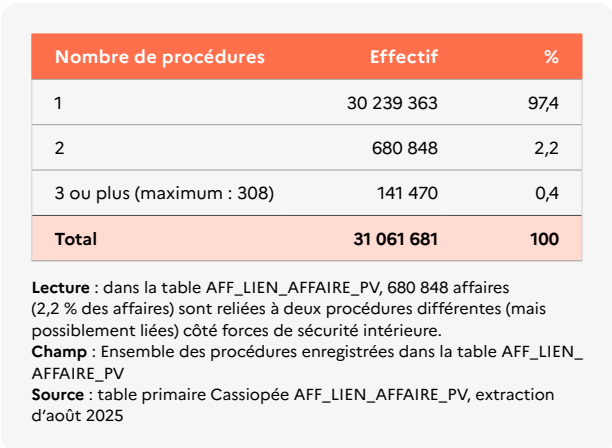
**Aperçu de la table primaire Cassiopée
AFF_LIEN_AFFAIRE_PV**

REFERENCE_PV_ID	AFFAIRE_ID	DATE_CREATION_LIEN	IDJ_DU_PV
28491XXX	100000000137965XXX	13/10/2022 22:04	220276XXXXG
28491XXX	100000000137965XXX	13/10/2022 22:02	220276XXXXF
28490XXX	100000000137964XXX	13/10/2022 21:01	220276XXXXE
28490XXX	100000000137964XXX	13/10/2022 21:02	220276XXXXR
28490XXX	100000000137964XXX	13/10/2022 21:02	220276XXXXF
28679XXX	100000000138222XXX	30/10/2022 22:02	220297XXXXF

• Figure 2. Nombre d’affaires par procédure



• Figure 3. Nombre de procédures par affaire



Chaque ligne de la table AFF_LIEN_AFFAIRE_PV correspond à un couple Affaire juridique – Procédure côté FSI. Le fichier comprend 32 149 000 lignes (à la date d’août 2025), soit la combinaison non nécessairement univoque de 30 515 000 procédures avec 31 062 000 affaires. Ainsi, dans les tables primaires de Cassiopée, 97 % des affaires juridiques correspondent à une seule procédure côté FSI, tandis que 96 % des procédures côté FSI correspondent à une affaire juridique unique. Le cas « 1 procédure – 1 affaire » est donc largement majoritaire.

Le rapprochement des deux tables primaires donne une clé d'appariement entre les deux systèmes d'information

Le rapprochement des deux tables AFF_LIEN_AFFAIRE_PV et ECH_REFERENTIEL_PV, grâce au numéro séquentiel [reference_pv_id], donne une correspondance (non univoque) entre l'identifiant des affaires côté justice et l'identifiant des procédures côté FSI, soit une clé d'appariement entre les données des deux ministères.

La table résultante de la jointure contient des triplets incomplets ou mal formatés. Pour les affaires provenant de la police ou de la gendarmerie, un triplet est considéré comme valide et utilisable pour l'appariement uniquement s'il remplit simultanément les quatre conditions suivantes :

- [annee_procedure] doit être comprise entre 2012 et 2025 (2012 correspond à l'année où Cassiopée a commencé à être déployée dans les juridictions),
- [code_unite_redactrice] est non nul, et de taille valide (5 ou 8 caractères selon que l'affaire est d'origine police ou gendarmerie),
- [numero_procedure] est non nul et de taille valide (5 ou 6 caractères selon que l'affaire est d'origine police ou gendarmerie),
- en outre, les tailles de [numero_procedure] et de [code_unite_redactrice] doivent être cohérentes entre elles (8 et 5, ou 5 et 6 pour le code service et le numéro de procédure selon que la procédure est d'origine police ou gendarmerie).

Au total, 96,7 % des affaires de la base AFF_LIEN_AFFAIRE_PV dont le PV est émis par la police ou la gendarmerie et enregistrées entre 2012 et 2024 sont associées à un triplet d'appariement « valide » (Figure 4). Les taux de validité sont de 94,7 % pour les données de la police nationale et 99,1 % pour les données de la gendarmerie nationale. La proportion d'affaires associées à un triplet valide varie selon les années : elle s'élève à plus de 98 % pour les affaires enregistrées en 2014-2015, puis chute de manière importante jusqu'en 2017 (91,9 %), avant de remonter. Pour les affaires enregistrées en 2024, 98,2 % des affaires juridiques sont associées à un triplet d'appariement « valide ».

Il est à noter que la moins bonne qualité des triplets constatée en 2016 et en 2017 est due à des problèmes techniques dans l'application Cassiopée sur cette période⁶. Ces problèmes ont été résorbés ensuite et il apparaît que la qualité des triplets s'améliore notablement sur les années récentes.

L'enrichissement par le Code statistique non signifiant (CSNS)

Le code statistique non signifiant ou CSNS (Figure 1) a été obtenu pour l'ensemble des mis en cause dans Cassiopée à la date de mai 2025⁷. Le fichier anonymisé des personnes mises en cause comprend près de 50 millions de lignes. 1,8 % lignes ont un CSNS vide (Figure 5). Quand le CSNS est rempli, la qualité de l'identification est correcte pour plus de 80 % d'entre eux, c'est-à-dire que leur indicateur de qualité est compris entre 1 (« parfaitement fiable ») et 4 (« assez fiable »). Les CSNS de très bonne qualité (indicateur de qualité égal à 1) représentent 59 % de l'ensemble.

• Figure 4. Proportion de triplets valides selon l'année d'enregistrement et l'origine de l'affaire (en %)

Année d'enregistrement	PN	GN	Ensemble
2012	42,0	88,0	85,2
2013	64,9	98,9	98,0
2014	95,7	99,3	98,6
2015	98,1	98,8	98,5
2016	95,6	97,4	96,4
2017	87,7	98,0	91,9
2018	91,7	98,7	94,5
2019	94,3	99,1	96,3
2020	94,9	99,3	96,9
2021	95,5	99,5	97,2
2022	95,4	99,6	97,3
2023	96,9	99,8	98,1
2024	97,1	99,8	98,2
Total	94,7	99,1	96,7

Lecture : après rapprochement avec la table ECH_REFERENTIEL_PV, 98,1% des affaires présentes dans la base AFF_LIEN_AFFAIRE_PV et enregistrées dans Cassiopée en 2023 sont associées à un triplet d'appariement « valide »
Champ : affaires enregistrées dans Cassiopée entre 2012 et 2024 et d'origine police (PN) ou gendarmerie (GN).
Source : ministère de la Justice, SSER, tables primaires Cassiopée, extractions d'août 2025.

• Figure 5. Répartition des mis en cause de Cassiopée selon qualité du CSNS

Note qualité	Nombre	%
1	29 567 327	59,0
2	8 229 427	16,4
3	1 316 584	2,6
4	300 250	0,6
5	1 921 274	3,8
6	782 162	1,6
7	7 041 287	14,1
Vide	924 894	1,8
Total	50 083 205	100,0

Lecture : 8 229 427 lignes du fichier des personnes mises en causes de Cassiopée (16,4 %) ont un CSNS de qualité 2.
Source : ministère de la Justice, SSER, Cassiopée (extraction de mai 2025) enrichi du CSNS.

6 Information relevée auprès de la Direction des services judiciaires du ministère de la Justice lors des échanges tenus régulièrement autour du projet d'appariement.
7 Les premières opérations d'enrichissement des données de Cassiopée par le CSNS ont été réalisées en 2024. Depuis la mi-2025, le SSER dispose d'un outil interne qui facilite les extractions.

Constitution des paquets de procédures dans les bases du SSMSI

L'unité statistique présente dans les données du SSMSI est la procédure. Une procédure est définie comme l'ensemble des actes d'enquête retranscrits sur des procès-verbaux. Chaque acte de la procédure fait l'objet d'une retranscription sur un procès-verbal distinct (audition, garde à vue, perquisition, par exemple). Lorsqu'il y a transfert de la procédure à un autre service, ou que l'enquête se poursuit, une nouvelle procédure peut être créée, alors qu'elle reste liée à un même dossier⁸ d'origine et sans qu'il soit toujours possible de chaîner les procédures liées. Il est donc possible que plusieurs procédures soient liées à une même affaire dans les données du SSER, bien qu'une seule procédure soit généralement référencée dans Cassiopée.

Il convient donc de regrouper *ex-post* les procédures des FSI pouvant être liées à une même affaire en justice. Pour ce faire, nous cherchons à créer des paquets de procédures en appliquant la méthode suivante :

- **Étape 1** : Création des couples de procédures potentiellement liées selon trois méthodes (utilisation de la procédure d'origine, de l'identité des victimes et de l'identité des mis en cause, via le CSNS).
- **Étape 2** : Validation des couples de procédures candidats de chaque méthode avec la commune de commission et la date des faits et élimination des faux positifs.
- **Étape 3** : Concaténation des couples validés à partir des trois méthodes.
- **Étape 4** : Regroupement des procédures en paquets à partir des couples validés via un algorithme de théorie des graphes.

Les paquets de procédures sont créés de façon indépendante pour les données de la police et de la gendarmerie nationales⁹:

- Pour la police nationale, le champ comprend l'ensemble des procédures créées entre 2018 et 2023, avec ou sans victime enregistrée, avec ou sans mis en cause enregistré.
- Pour la gendarmerie nationale, le champ comprend les procédures contenant la première constatation et/ou la première élucidation, créées entre 2018 et 2023, avec ou sans victime enregistrée, avec ou sans mis en cause enregistré.

Création des couples de procédures avec la procédure d'origine

Dans LRPPN et LRPGN, une variable est disponible renseignant la procédure d'origine, lorsqu'il en existe effectivement une : cette variable permet de faire le lien entre la procédure d'intérêt et une procédure antérieure qui lui serait liée. Les couples de procédures sont créés en utilisant le numéro de la procédure d'intérêt et le numéro de la procédure d'origine renseignée.

Le numéro de procédure d'origine est renseigné dans le cadre d'une transmission de procédure à un autre service. Il est rempli à la main dans le logiciel de rédaction, il faut donc s'assurer de sa validité. Lorsque le numéro de procédure d'origine n'est pas retrouvé dans les numéros de procédures

présents dans les puits de données, le numéro de procédure d'origine renseigné est considéré comme faux et est effacé. 46,1 % des couples de procédures PN (police nationale) et 20,5 % des couples de procédures GN (gendarmerie nationale) créés via cette méthode sont ainsi supprimés.

Les couples de procédures ainsi formés sont validés de façon séquentielle : dans un premier temps, il est nécessaire qu'il existe au moins un fait dans les deux procédures du couple dont la commune de commission est identique puis, dans un second temps, il est nécessaire que l'écart entre les dates de début de ce fait soit inférieur ou égal à 1 jour (**Figure 6** et **Figure 7**).

Pour la police nationale :

- Il y a 21 686 000 procédures présentes dans les puits de données PN entre 2018 et 2023.
- La méthode d'utilisation de la variable de procédure d'origine (après effacement des procédures d'origine considérées fausses) permet de créer 972 000 couples de procédures candidats avant validation.
- 663 000 couples sont validés par la commune et la date des faits, soit 68,2 % des couples candidats.

• Figure 6. Répartition des couples de procédures de la police nationale selon la concordance de la commune de commission et de la date des faits (en %)

	Dates non valides (écart > 1j)	Dates valides (écart <= 1j)
Communes différentes	16,8	-
Communes identiques	15,0	68,2

*La validité de la date n'est testée que lorsque les communes de commission sont bien identiques.

Source : extraction du puits LRPPN - SSMSI.

Pour la gendarmerie nationale :

- Il y a 7 661 000 procédures présentes dans les puits de données GN entre 2018 et 2023.
- La méthode d'utilisation de la variable de procédure d'origine (après effacement des procédures d'origine considérées fausses) permet de créer 348 000 couples de procédures candidats avant validation.
- 278 000 couples sont validés par la commune de commission et la date des faits, soit 79,9 % des couples candidats.

• Figure 7. Répartition des couples de procédures de la gendarmerie nationale selon la concordance de la commune de commission et de la date des faits (en %)

	Dates non valides (Écart > 1j)	Dates valides (Écart <= 1j)
Communes différentes	10,5	-
Communes identiques	9,6	79,9

*La validité de la date n'est testée que lorsque les communes de commission sont bien identiques.

Source : extraction du puits MIS - SSMSI.

⁸ Le terme « dossier » utilisé ne correspond pas à une unité statistique existante dans les données du SSMSI : il est utilisé uniquement dans le cadre de cette étude pour désigner plusieurs procédures liées sans parler d'affaire à ce stade, qui pourrait être confondue avec l'unité statistique des données du SSER. L'objectif est que le dossier se rapproche le plus possible de l'affaire.

⁹ Il s'agit d'une contrainte de production, il existe formellement des transferts de procédures entre la police et la gendarmerie nationales.

Création des couples de procédures avec l'identité des victimes et des mis en cause

Les identifiants de victime et les identifiants de mis en cause présents dans les puits de données sont propres à chaque procédure. Il n'est a priori pas possible, à partir de ces identifiants, de repérer si un individu a été victime ou mis en cause dans plusieurs procédures différentes. C'est pourquoi le CSNS est utilisé pour repérer les victimes et/ou mis en cause communs à plusieurs procédures.

Des couples de procédures sont créés en rapprochant deux procédures qui ont au moins une victime (ou un mis en cause) avec un CSNS commun entre 2018 et 2023. Cependant, le fait d'être victime (ou mis en cause) dans plusieurs procédures ne permet pas d'affirmer que deux procédures sont liées : il est possible d'être victime (ou mis en cause) plusieurs fois pour des dossiers distincts.

Pour pallier cela, et comme pour la méthode précédente, un couple de procédures est validé s'il existe au moins un fait dans les deux procédures du couple dont la commune de commission est identique et l'écart entre les dates de début des faits est inférieur ou égal à 1 jour.

Des couples sont créés indépendamment pour les données de la police nationale et la gendarmerie nationale en utilisant le CSNS des victimes d'une part, et le CSNS des mis en cause d'autre part. Cette méthode reste complémentaire à celle présentée dans la partie précédente dans la mesure où il n'y a pas toujours de victime et/ou de mis en cause dans les procédures.

Pour la police nationale :

- Il y a 21 686 000 procédures présentes dans le puits de données PN entre 2018 et 2023.
- La méthode d'utilisation du CSNS des victimes permet de créer 9 101 000 couples de procédures candidats avant validation.
- 241 000 couples sont validés par la date et le lieu des faits, soit 2,6 % des couples candidats.
- La méthode d'utilisation du CSNS des mis en cause permet de créer 7 739 000 couples de procédures candidats avant validation.
- 79 000 couples sont validés par la date et le lieu des faits, soit 1,0 % des couples candidats.

Pour la gendarmerie nationale :

- Il y a 7 661 000 procédures présentes dans le puits de données GN entre 2018 et 2023.
- La méthode d'utilisation du CSNS des victimes permet de créer 4 361 000 couples de procédures candidats avant validation.
- 103 000 couples sont validés par la date et le lieu des faits, soit 2,4 % des couples candidats.
- La méthode d'utilisation du CSNS des mis en cause permet de créer 4 506 000 couples de procédures candidats avant validation.
- 108 000 couples sont validés par la date et le lieu des faits, soit 2,4 % des couples candidats.

Cette méthode de rapprochement de procédures via l'identité des victimes et des mis en cause ne prend pas en compte à ce stade le champ infractionnel, ce qui pourrait permettre d'avoir des rapprochements de meilleure qualité (arbitrage entre faux positifs et faux négatifs à étudier).

Regroupement des méthodes et création des paquets de procédures

Les trois méthodes précédemment présentées sont concaténées indépendamment pour la police nationale et la gendarmerie nationale afin de regrouper tous les couples de procédures valides. Elles permettent de détecter 953 000 couples de procédures pour la police nationale, dont 3,2 % des couples sont détectés par au moins deux méthodes, et 457 000 couples pour la gendarmerie nationale, dont 7,0 % sont détectés par au moins deux méthodes.

La théorie des graphes (Figure 8) est ensuite utilisée pour regrouper l'ensemble des couples de procédures précédemment créés.

• Figure 8. Théorie des graphes pour regrouper les procédures

En disposant de couples de procédures, il est possible de modéliser l'ensemble des relations sous forme d'un graphe. Dans ce graphe :

- chaque nœud représente une procédure,
- chaque arête relie deux procédures associées dans un couple validé.

Ainsi, les couples de procédures deviennent une structure de graphe non orienté, où la connexité des nœuds traduit les dépendances ou les relations entre les procédures.

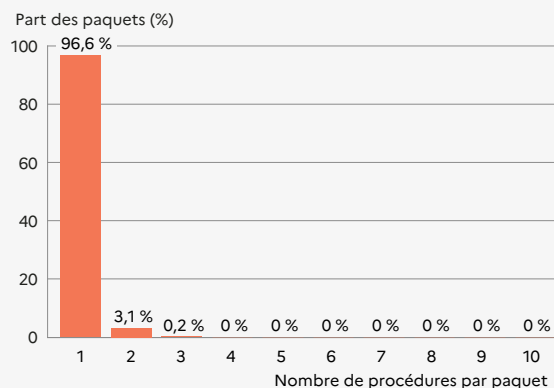
L'objectif est ensuite d'identifier les groupes cohérents de procédures, c'est-à-dire des ensembles au sein desquels toutes les procédures sont directement ou indirectement reliées entre elles : ces groupes correspondent aux composantes connexes du graphe.

Une composante connexe est un sous-ensemble maximal de nœuds tels que chaque paire de nœuds est reliée par au moins un chemin. Autrement dit, si deux procédures appartiennent à la même composante connexe, il existe une chaîne de relations (directes ou via d'autres procédures) qui les relie. Ces composantes sont identifiées par un algorithme qui parcourt en profondeur le graphe (*Depth-First Search, DFS*).

Chaque composante connexe correspond alors à un paquet de procédures.

La plupart des procédures sont isolées dans le graphe dans le sens où elles ne sont reliées à aucune autre procédure, comme présenté dans les [Figure 9](#) et [Figure 10](#).

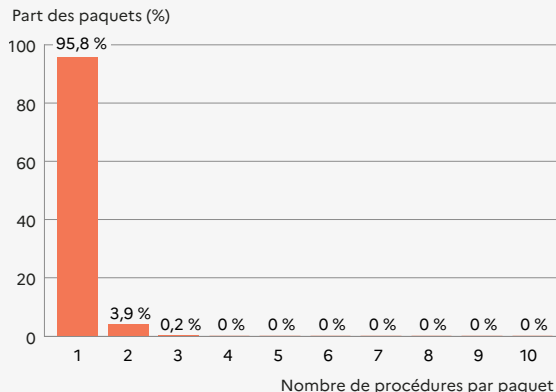
• **Figure 9. Distribution de la taille des paquets de procédure de la police nationale**



Note de lecture : 96,6 % des paquets PN sont constitués d'une seule procédure.

Source : puits de données LRPPN 2018-2023.

• **Figure 10. Distribution de la taille des paquets de procédure de la gendarmerie nationale**



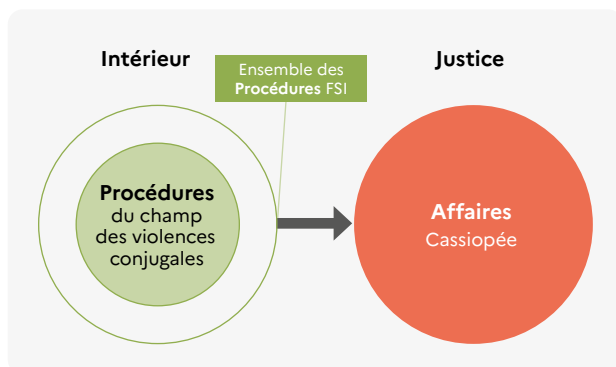
Lecture : 95,8 % des paquets GN sont constitués d'une seule procédure.

Source : puits de données MIS 2018-2023.

Appariement sur caractéristiques communes sur le champ des violences conjugales

L'appariement des données entre le ministère de l'Intérieur et le ministère de la Justice vise prioritairement à suivre le devenir judiciaire des plaintes enregistrées par les forces de sécurité intérieure. Les procédures enregistrées par les FSI sont donc le point de départ de l'appariement, pour lesquelles sont recherchées les affaires judiciaires correspondantes dans Cassiopée. Par souci de soutenabilité computationnelle¹⁰, il est nécessaire de travailler par champ infractionnel : dans la suite de ce document, nous nous concentrons sur le champ des violences conjugales que nous cherchons à appairer avec l'ensemble des affaires enregistrées dans Cassiopée ([Figure 11](#) et [Figure 12](#)).

• **Figure 11. Schéma représentatif de l'appariement : de la procédure à la réponse pénale**



L'appariement entre les procédures enregistrées par les FSI et les affaires Cassiopée s'effectue en plusieurs étapes. Cela implique une mise en cohérence d'unités statistiques distinctes (les procédures côté intérieur et les affaires côté justice) ainsi que le recours à deux méthodes

¹⁰ Une base statistique annuelle sur les victimes enregistrées par les FSI contient environ 3,5 millions de victimes et les données Cassiopée environ 31 millions d'affaires.

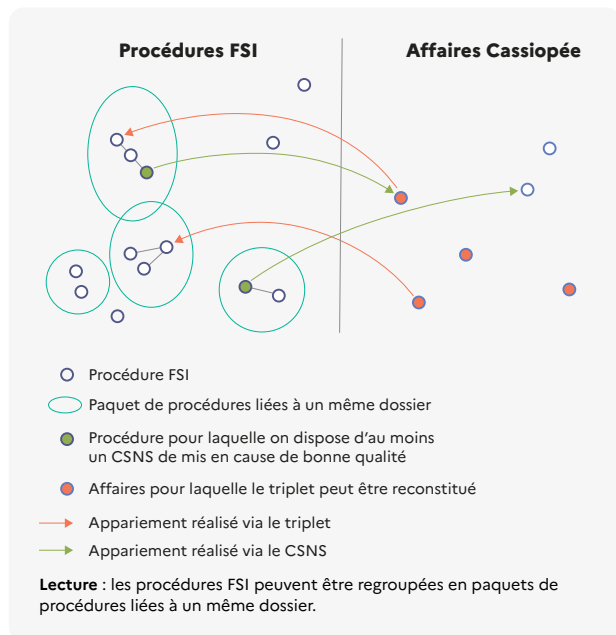
• **Figure 12. Définition du champ des violences conjugales dans les données issues des FSI**

Le périmètre infractionnel est défini à partir de la nomenclature française des infractions (NFI), selon les catégories d'infractions suivantes :

- « **tentatives d'homicide** » : 01.A (homicides intentionnels) / 01.C1 (atteintes volontaires ayant entraîné la mort sans intention de la donner) ; les tentatives d'homicide se distinguent des homicides aboutis par leurs index (classification agrégée des crimes et délits en 107 postes, dont disposent les services de police et de gendarmerie, et utilisée historiquement par le ministère de l'Intérieur pour établir les statistiques de la délinquance).
- « **violences physiques** » : 02.A1 (tortures), 02.A2 (violences) / 02.A9 (autres violences ou menaces) / 02.F6 (administration de substance nuisible)
- « **harcèlement moral** » : 02.G1 (harcèlement moral au travail) / 02.G2 (harcèlement moral sur conjoint) / 02.G3 (harcèlement moral)
- « **menaces** » : 02.A4 (menaces) / 02.G7 (menaces d'atteintes aux biens) / 02.E (extorsion ou chantage)
- « **autres violences psychologiques ou verbales** » : 02.G4 (appels téléphoniques malveillants) / 02.G5 (usurpation d'identité pour nuire à l'honneur ou à la considération de la personne) / 02.H (diffamation ou injure) / 02.J (atteintes à l'intimité de la personne)
- « **viol et tentative de viol** » : 03.A (viols)
- « **autres violences sexuelles** » : 03.B (agressions ou atteintes sexuelles) / 03.C (violences sexuelles non physiques) / 03.D (exploitation sexuelle)

d'appariement complémentaires : une utilisant le triplet (année d'enregistrement, code service, numéro de procédure) et l'autre le CSNS des mis en cause. Les relations entre les procédures, les paquets de procédures et les affaires sont multiples et non univoques : une affaire peut être liée à plusieurs procédures (ou paquets de procédures) et, à l'inverse, une procédure (ou paquet de procédures) peut être liée à plusieurs affaires. La **Figure 13** représente la complexité de l'appariement.

• **Figure 13. Représentation de l'appariement entre les procédures et les affaires**



Il est possible qu'une procédure de violences conjugales appartienne à un paquet de procédures dans lequel une autre procédure n'appartient pas au champ des violences conjugales tel que défini dans la **Figure 13**. Le champ des procédures de violences conjugales originel est donc élargi à l'ensemble des procédures appartenant à un paquet de procédures dans lequel au moins une procédure appartient au champ des violences conjugales.

Cette partie présente dans un premier temps la méthodologie d'appariement via le triplet, ainsi que la validation de cet appariement par date et lieu des faits. Dans un second temps, les mêmes éléments sont présentés pour l'appariement via le CSNS des mis en cause, avant de regrouper les deux méthodes d'appariement et d'enrichir les taux d'appariement avec les paquets de procédures préalablement créés. La dernière partie analysera les relations multiples entre les procédures et les affaires.

Appariement sur triplet et validation des appariements

La clé d'appariement principale est une variable construite comme un triplet de variables présentes dans les deux systèmes d'information (Année de la création de la procédure / Code de l'unité rédactrice de la procédure/ Numéro de la procédure). L'ordre de concaténation varie selon l'origine de la procédure (police nationale ou gendarmerie nationale) (**Figure 14**). Il s'agit de l'identifiant créé par les logiciels de rédaction de procédure. Ce triplet est en principe transmis à la Justice mais en réalité de manière imparfaite (**Figure 15**).

• **Figure 14. Constitution de la clé d'appariement selon l'origine de la procédure (FSI)**



• **Figure 15. La qualité du triplet dans Cassiopée pour le champ des violences conjugales**

Le triplet est renseigné de manière imparfaite dans Cassiopée. Si ce dernier est systématiquement rempli dans les données du ministère de l'Intérieur en tant que numéro unique d'enregistrement, son inscription dans Cassiopée nécessite souvent un remplissage manuel. Les critères de validité du triplet sont exposés en détail dans la section « Données du service statistique du ministère de la Justice (SSER) » (voir *supra*).

Le triplet est en général très bien renseigné pour les affaires ayant une nature d'affaire « Violences par conjoint ou concubin » (Nataff A38), avec une moyenne de 96,8 % de triplets valides pour les affaires enregistrées entre 2018 et 2023. Il existe des disparités sur la qualité du remplissage selon le champ infractionnel considéré. Les homicides, par exemple, présentent des taux de validité particulièrement dégradés, s'expliquant notamment par un remplissage moins important du code service. Le remplissage de la clé d'appariement est de meilleure qualité pour les affaires d'origine gendarmerie nationale que pour celles d'origine police nationale.

70,9 % des procédures du champ des violences conjugales issues de la base statistique Victimes 2023, tous services (TS) compris, sont appariées avec une affaire Cassiopée par le triplet (**Figure 16**). Il existe des disparités entre la police nationale et la gendarmerie nationale : 65,8 % des procédures du champ précédent enregistrées par la police nationale sont appariées contre 78,6 % pour celles enregistrées par la gendarmerie nationale. Cette disparité est difficile à expliquer malgré des triplets mieux renseignés côté Cassiopée pour les procédures gendarmerie que pour celles issues de la police.

Le taux d'appariement est plus faible au cours des années récentes, notamment à partir de 2022. Cela est dû au manque de recul temporel provenant du délai entre l'enregistrement de la procédure par les FSI et la saisie de l'affaire associée dans Cassiopée.

• **Figure 16. Taux d'appariement via le triplet des procédures de violences conjugales avec au moins une affaire Cassiopée entre 2018 et 2023**

Année	Nb proc. PN	Nb proc. GN	Nb proc. TS	Taux PN (%)	Taux GN (%)	Taux TS (%)
2018	75 326	46 285	121 611	71,8	80,3	75,1
2019	85 052	54 958	140 010	71,1	80,4	74,7
2020	88 903	63 057	151 960	71,1	81,6	75,5
2021	106 756	70 160	176 916	70,9	81,3	75,0
2022	122 564	78 105	200 669	68,9	80,2	73,3
2023	131 963	86 560	218 523	65,8	78,6	70,9

Note de lecture : 70,9 % des procédures enregistrées en 2023 par un service de police ou de gendarmerie nationale sont appariées avec au moins une affaire Cassiopée par le triplet.
Source : extraction Cassiopée août 2025 ; bases statistiques Victimes du SSMSI 2018 - 2023.

Un appariement par triplet est considéré comme valide si les deux identifiants correspondent dans les données du SSMSI et celles du SSER, sans contrôle sur une autre variable. Un contrôle de la qualité de l'appariement peut être effectué sur des variables annexes : la date et le lieu des faits. Parmi les appariements réalisés par le triplet, 80 % des couples procédure-affaire ont un écart entre la date des faits enregistrés par les FSI et celle saisie dans Cassiopée inférieur ou égal à cinq jours. La présence d'écarts de date ne signifie pas pour autant que ces appariements sont erronés : il peut y avoir de légères erreurs dans l'enregistrement de la date de fait de part et d'autre de la chaîne pénale.

Appariement sur CSNS et validation des appariements

Au-delà du triplet comme identifiant commun, les procédures et les affaires enregistrées contiennent

d'autres informations permettant de les rapprocher : les informations d'état civil des mis en cause (nom, prénom, date de naissance, sexe) via le CSNS (Figure 1). Le CSNS des mis en cause est utilisé pour réaliser des appariements en complément de la méthode d'appariement par le triplet. Toutefois, le CSNS seul ne suffit pas à rapprocher une procédure et une affaire. Il est en effet possible pour un individu d'être mis en cause dans plusieurs procédures/affaires distinctes. Il faut donc utiliser en complément des variables de validation pour s'assurer de la pertinence de l'appariement : la date et la commune des faits.

La réussite de l'appariement de deux CSNS présents dans les bases respectives de chaque SSM dépend de la qualité de ces CSNS : il est nécessaire que les informations d'état civil des mis en cause/auteurs soient suffisamment bien renseignées dans chaque système d'information.

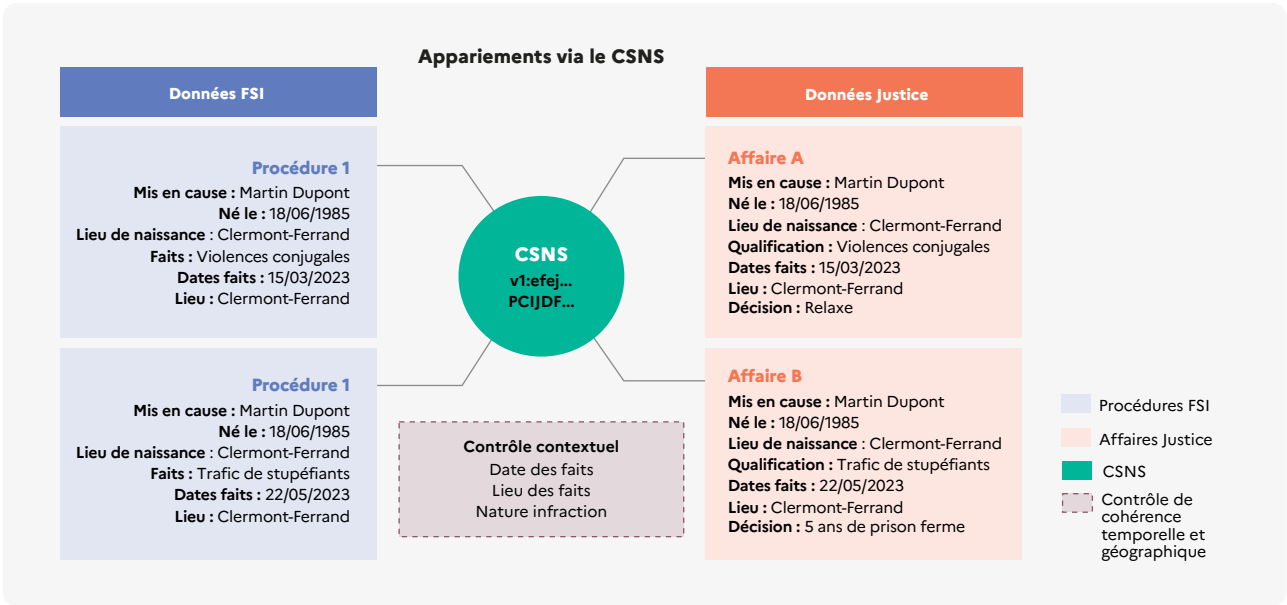
Le rapprochement des bases via le CSNS des mis en causes peut être non univoque. La Figure 17 présente un exemple dans lequel le rapprochement via le CSNS seul génère 4 couples procédures-affaires : Martin Dupont est mis en cause dans plusieurs dossiers, or les couples *Procédure 1 - Affaire B* et *Procédure 2 - Affaire A* ne correspondent pas au même dossier et ne doivent pas être appariés.

Ainsi il est nécessaire de contrôler les couples de procédures-affaires créés par le CSNS par la date des faits et la commune de commission des faits pour pouvoir *in fine* effectivement rattacher la procédure à l'affaire qui lui correspond. La règle de validation d'un appariement procédure - affaire via le CSNS est la suivante :

- Écart de date de fait inférieur ou égal à 1 jour,
- Correspondance exacte de la commune de commission.

Chaque couple est validé s'il existe au moins une infraction¹¹ dans la procédure pour laquelle la règle précédente est respectée. Pour l'heure, il n'y a pas de contrôle effectué sur la qualification de l'infraction, notamment du fait des différences de nomenclature, voire de classification dans les deux systèmes d'information.

• **Figure 17. Exemple d'appariement via le CSNS**



¹¹ Une procédure peut comprendre plusieurs infractions.

• **Figure 18. Taux d'appariement via le CSNS des mis en cause des procédures de violences conjugales avec au moins une affaire Cassiopée entre 2018 et 2023**

Année	Nb proc. PN	Nb proc. GN	Nb proc. TS	Taux PN (%)	Taux GN (%)	Taux TS (%)
2018	75 326	46 285	121 611	14,2	48,7	27,3
2019	85 052	54 958	140 010	36,5	47,6	40,9
2020	88 903	63 057	151 960	33,2	51,2	43,6
2021	106 756	70 160	176 916	40,2	50,2	44,2
2022	122 564	78 105	200 669	40	48,6	43,4
2023	131 963	86 560	218 523	37,9	48,1	41,9

Note de lecture : 41,9 % des procédures enregistrées en 2023 par un service de police ou de gendarmerie nationales sont appariées avec au moins une affaire Cassiopée par le CSNS des mis en cause.
Source : extraction Cassiopée août 2025 ; bases statistiques Victimes du SSMSI 2018 - 2023 ; extraction du puits SSMSI.

41,9 % des procédures du champ des violences conjugales issues de la base statistique Victimes 2023, tous services (TS) compris, sont appariées par le CSNS des mis en cause (Figure 18). Le taux d'appariement est beaucoup plus faible pour l'année 2018 en police nationale : cela s'explique par une mauvaise qualité des données nominatives, la durée de conservation de ces données étant de 5 ans¹². De plus, les CSNS des mis en cause sont de meilleure qualité pour les procédures enregistrées par la gendarmerie nationale, ce qui explique le meilleur taux d'appariement en gendarmerie nationale comparé à la police nationale. Les taux d'appariement via le CSNS des mis en cause sont inférieurs par rapport à ceux observés via le triplet, toutes les procédures n'ayant pas de mis en cause. L'intérêt de cette seconde méthode d'appariement est de compléter l'appariement via le triplet préalablement présenté. Il convient donc d'analyser si les procédures et affaires appariées via cette méthode le sont également via le triplet ou non, et de mesurer l'apport de cet appariement par rapport à la première méthode.

Regroupement des deux méthodes d'appariement

Chaque procédure peut être appariée à une ou plusieurs affaires par une seule des méthodes ou les deux à la fois. Par conséquent, il convient d'analyser l'apport de chaque méthode pour l'appariement sur le champ des violences conjugales. L'appariement par le triplet permet l'apport le plus important en termes de couverture et demeure donc la clé d'appariement principale. L'appariement via le CSNS des mis en cause permet, sur le champ des violences conjugales, d'accroître la couverture du nombre de procédures appariées de +2,1 points de pourcentage au minimum en 2018 à +4,8 points de pourcentage au maximum en 2022 par rapport à l'utilisation seule du triplet comme clé d'appariement (Figure 19). Une large part des procédures appariées via le CSNS des mis en cause l'est donc déjà par le triplet.

Au final, au plus bas, 74,9 % des procédures de violences conjugales sont appariées avec au moins une affaire Cassiopée en 2023 et, au plus haut, 79,0 % des procédures de violences conjugales sont appariées avec au moins une affaire Cassiopée en 2020 et en 2021.

• **Figure 19. Taux d'appariement via le triplet et le CSNS des mis en cause des procédures de violences conjugales avec au moins une affaire Cassiopée entre 2018 et 2023 (en %)**

Année	Triplet uniquement			Triplet et CSNS		
	Taux PN	Taux GN	Taux TS	Taux PN	Taux GN	Taux TS
2018	71,8	80,3	75,1	73,2	83,7	77,2
2019	71,1	80,4	74,7	75,0	83,8	78,5
2020	71,1	81,6	75,5	75,2	85,1	79,0
2021	70,9	81,3	75,0	75,2	84,8	79,0
2022	68,9	80,2	73,3	74,1	84,4	78,1
2023	65,8	78,6	70,9	70,2	82,0	74,9

Note de lecture : en 2023, 70,9 % des procédures TS sont appariées via le triplet uniquement, contre 74,9 % en combinant les deux méthodes d'appariement via le triplet et le CSNS des mis en cause.

Source : extraction Cassiopée août 2025 ; bases statistiques Victimes du SSMSI 2018 - 2023 ; extraction du puits SSMSI.

Apport des paquets de procédures pour l'appariement

Environ un cinquième des procédures de violences conjugales constatées demeurent non appariées à au moins une affaire Cassiopée. Cependant, pour une même affaire judiciaire de violences conjugales, plusieurs procédures peuvent avoir été enregistrées par les FSI (cf. supra la partie « Constitution des paquets de procédures dans les bases du SSMSI »). Dans un tel cas de figure, ces procédures sont liées et il est possible qu'une seule de ces procédures soit présente dans les données Cassiopée. Un changement d'unité statistique est donc pertinent pour analyser les taux d'appariements. Cette partie présente donc les taux d'appariements des paquets de procédures de violences conjugales avec au moins une affaire Cassiopée. Un paquet de procédures est considéré comme apparié si au moins une procédure appartenant à ce paquet est appariée à au moins une affaire Cassiopée.

Les procédures constatées de violences conjugales peuvent appartenir à un paquet de procédures dans lequel certaines procédures n'appartiennent pas au champ originel des violences conjugales. Ces procédures, hors champ des violences conjugales mais appartenant à un paquet de procédures dans lequel au moins une autre procédure est de violence conjugale, peuvent aussi être appariées à une affaire Cassiopée : elles sont donc ajoutées à l'analyse. De même, les paquets de procédures peuvent contenir des procédures appartenant au champ des violences conjugales mais qui ne correspondent pas à la procédure de premier enregistrement de l'infraction (commission rogatoire par exemple). Ces procédures, qui ne sont pas dans les bases statistiques du SSMSI (cf. Partie « Données du service statistique du ministère de l'Intérieur (SSMSI) »), peuvent aussi être appariées à au moins une affaire Cassiopée : elles sont donc également ajoutées à l'analyse.

En résumé, les procédures ajoutées à l'analyse sont :

- les procédures non constatées contenant au moins une infraction de violences conjugales, appartenant à un paquet de procédures ayant au moins une procédure de constatation du champ des violences conjugales,

¹² Les CSNS des mis en cause ont été produits en 2024.

- les procédures hors champ des violences conjugales (constatées ou non) appartenant à un paquet de procédures dans lequel il y a au moins une procédure de constatation de violences conjugales.

Les taux d'appariements sont naturellement plus élevés en utilisant l'unité statistique du paquet de procédures plutôt que de la procédure (Figure 20). Cela s'explique par deux phénomènes :

- D'une part, plusieurs procédures du champ initial peuvent appartenir à un même paquet dont une seule procédure est appariée (effet baisse du dénominateur).
- D'autre part, certaines procédures hors du champ des procédures de constatation de violences conjugales peuvent être appariées alors qu'aucune autre procédure du champ des violences conjugales appartenant à ce paquet n'est appariée (effet hausse du numérateur).

Au final, au plus bas, 77,4 % des paquets de procédures de violences conjugales sont appariées avec au moins une affaire Cassiopée en 2023 et, au plus haut, 82,1 % des paquets de procédures de violences conjugales sont appariées avec au moins une affaire Cassiopée en 2020. L'apport de la constitution des paquets de procédure est donc de l'ordre de 2 à 3 points de pourcentage sur le champ des violences conjugales.

• **Figure 20. Comparaison des taux d'appariement avec au moins une affaire Cassiopée des procédures de violences conjugales et de celui des paquets de procédures conjugales entre 2018 et 2023 (en %)**

Année	Niveau Procédure			Niveau Paquet		
	PN	GN	TS	PN	GN	TS
2018	73,2	83,7	77,2	74,1	86,8	78,9
2019	75,0	83,8	78,5	76,9	86,8	80,8
2020	75,2	85,1	79,0	78,1	87,7	82,1
2021	75,2	84,8	79,0	78,1	87,4	81,8
2022	74,1	84,4	78,1	76,6	87,1	80,7
2023	70,3	82,0	74,9	72,3	85,1	77,4

Note de lecture : en 2022, tous services confondus, 78,1 % des procédures constatées de violences conjugales sont appariées avec au moins une affaire Cassiopée et 80,7 % des paquets de procédures de violences conjugales sont appariés avec au moins une affaire Cassiopée.
Champ : procédures constatées des violences conjugales appariées avec l'ensemble affaires Cassiopée et paquets de procédures issues du champ des violences conjugales appariées avec l'ensemble affaires Cassiopée.
Source : extraction Cassiopée août 2025 ; bases statistiques Victimes du SSMSI 2018 - 2023 ; extraction du puits SSMSI.

Les taux d'appariement de la Figure 20 sont calculés comme suit :

$$\text{Taux d'appariement}_{\text{procédure}} = \frac{\text{Nombre de procédures appariées}}{\text{Nombre de procédures total}} \times 100$$

$$\text{Taux d'appariement}_{\text{paquet}} = \frac{\text{Nombre de paquets appariés}}{\text{Nombre de paquets total}} \times 100$$

Analyse des relations multiples entre les procédures et les affaires

La méthodologie d'appariement engendre des relations multiples entre les procédures et les affaires : une ou plusieurs procédures peuvent être appariées à une ou plusieurs affaires. L'analyse de ces relations multiples permet d'une part d'améliorer la compréhension des relations entre les procédures et les affaires. D'autre part,

elle permet de mettre en lumière certaines incohérences du processus d'appariement, par exemple lorsque plusieurs affaires appariées à une même procédure paraissent ne pas être liées.

La Figure 21 présente une liste des relations multiples possibles et les enseignements que chacune d'elles permettent de tirer.

• **Figure 21. Relations multiples possibles**

Une procédure - Plusieurs affaires :

- Il peut s'agir d'une procédure impliquant un mineur et un majeur qui peut générer deux affaires distinctes côté Justice, ou, plus rarement d'une procédure avec plusieurs mis en cause, faisant l'objet d'orientations différentes selon leur degré d'implication dans les infractions poursuivies.
- Il peut s'agir d'enregistrement d'affaires en doublons dans Cassiopée, notamment en lien avec des jonctions d'affaires.

Plusieurs procédures - Une affaire

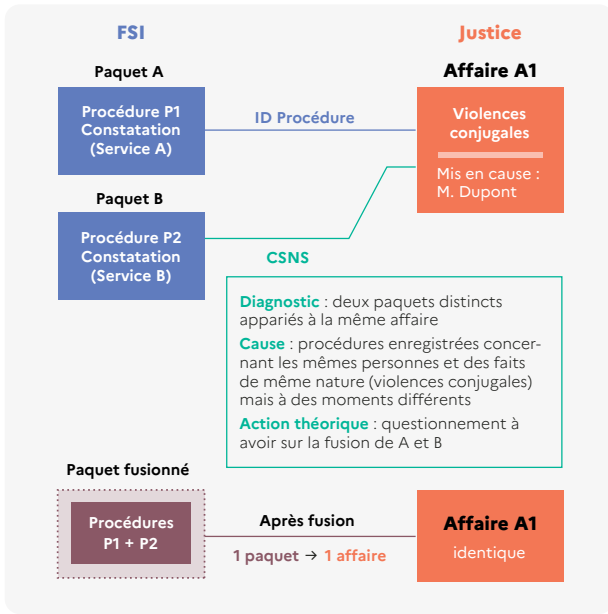
- Plusieurs procédures peuvent alimenter une même affaire judiciaire, c'est dans ce cadre que sont créés les paquets de procédures pour réconcilier les unités statistiques (exemple : procédure initiale + procédures d'enquête complémentaire qui sont appariées à une affaire unique).

Plusieurs procédures - Plusieurs affaires

- Plusieurs procédures de paquets de procédures différents peuvent être appariées à une même affaire, ce qui peut suggérer que ces paquets de procédures devraient être joints.
- Plusieurs procédures d'un même paquet de procédures peuvent être appariées à des affaires différentes, ce qui peut suggérer un mauvais appariement.

La Figure 22 présente un cas d'exemple de deux procédures appariées à une même affaire : une procédure est appariée par le triplet et l'autre par le CSNS des mis en cause.

• **Figure 22. Exemple d'une relation de deux procédures de deux paquets différents appariées à une même affaire**



Or ces deux procédures n'appartiennent pas au même paquet de procédures. Cela s'explique par le fait qu'elles correspondent toutes deux à des faits de violences conjugales qui se sont déroulés à des moments différents. Toutefois, pour le parquet, il n'y a qu'une seule et même affaire enregistrée dans Cassiopée qui contient l'ensemble des faits. Si une personne est mise en cause pour plusieurs faits / procédures distinctes, le parquet peut choisir de la juger en une seule fois (notamment s'il y a une certaine cohérence des faits poursuivis).

La **Figure 23** présente un exemple de deux affaires différentes qui sont appariées à une même procédure. Les dates des faits des deux affaires sont séparées de 19 jours, la première affaire est classée au mois de mars alors que la seconde est classée en octobre. Les dates d'enregistrements au parquet sont également différentes pour les deux affaires. Seul le tribunal de Grande Instance est identique entre les deux affaires. Cela tend à montrer que ces deux affaires ne sont en réalité pas liées et qu'une seule des deux affaires devrait être appariée à la procédure.

• **Figure 23. Exemple de cas de deux affaires appariées à la même procédure**

ID affaire	Date faits	Événement	Date événement	TGI	Date enregist. parquet
52252XXX	2021-02-06	PVSAI	2021-02-07	TGI/BOBIGNY	2021-02-24
52252XXX	2021-02-06	CLASSMT	2021-12-20	TGI/BOBIGNY	2021-02-24
52252XXX	2021-02-06	CLASSMT	2022-03-16	TGI/BOBIGNY	2021-02-24
52252XXX	2021-01-25	PVSAI	2022-08-22	TGI/BOBIGNY	2022-10-13
52252XXX	2021-01-25	CLASSMT	2022-10-13	TGI/BOBIGNY	2022-10-13
52252XXX	2021-01-25	CLASSMT	2022-10-13	TGI/BOBIGNY	2022-10-13

Source : données Cassiopée

Caractérisation des procédures appariées et non-appariées sur le champ des violences conjugales

L'objectif de cette partie est de déterminer, sur le champ des violences conjugales, s'il existe des caractéristiques spécifiques aux procédures non appariées par rapport aux procédures appariées. Ces caractéristiques proviennent des autres variables caractérisant les procédures qui sont présentes dans les puits de données du ministère de l'Intérieur.

Par simplicité, nous nous restreignons aux procédures de violences conjugales admettant au moins une victime en 2022. Sur cette année, sans prendre en compte les paquets de procédures, 78,1 % des procédures sont appariées (74,1 % en police nationale et 84,4 % en gendarmerie nationale) (**Figure 19**).

L'unité statistique appariée étant la procédure, les variables annexes utilisées pour caractériser l'appariement sont uniquement celles définies au niveau de la procédure (les variables caractéristiques des victimes et des mis en cause ne sont donc pas utilisées).

Dans un premier temps, nous cherchons à conserver les variables influentes sur l'appariement de la procédure via une analyse des correspondances multiples (ACM). Ensuite, nous analysons l'impact direct des variables les plus influentes sur le taux d'appariement.

Identification des variables par analyse des correspondances multiples (ACM)

Une analyse des correspondances multiples (**Figure 24**) est effectuée indépendamment pour les données de la police nationale d'une part et de la gendarmerie nationale d'autre part. La variable d'intérêt (le fait que la procédure soit appariée ou non) est mise en variable supplémentaire pour ne pas qu'elle influence la construction des axes factoriels.

• **Figure 24. Analyse des correspondances multiples (ACM)**

L'Analyse des Correspondances Multiples (ACM) est une méthode qui permet d'étudier les relations entre plusieurs variables qualitatives. Elle sert à créer des résumés des caractéristiques des procédures PN et GN qui contiennent la plus grande partie de l'information contenue dans ces variables. Ces résumés, appelés axes de l'ACM, sont constitués pour que le premier axe corresponde à la combinaison des caractéristiques des procédures contenant la plus grande partie de l'information du nuage de point initial. Le deuxième axe est nécessairement orthogonal au premier axe et forme donc un résumé indépendant de l'information du nuage de points qui n'aurait pas été prise en compte dans le premier axe, et de même pour les axes suivants.

Une fois ces axes constitués, il est ensuite possible de les comparer à une variable indicatrice signalant si la procédure a été appariée à une affaire en justice ou non (variable supplémentaire). Ainsi, l'ACM met en évidence les profils typiques de procédures les plus fortement associés à l'appariement : les modalités proches de la modalité « appariée » dans le plan factoriel partagent des caractéristiques communes qui augmentent la probabilité d'appariement, tandis que celles éloignées s'y associent peu. Cette approche permet ainsi une lecture synthétique et visuelle des correspondances structurelles entre les variables descriptives et le statut d'appariement des procédures, nous permettant ainsi de faire ressortir les variables les plus influentes parmi celles analysées sur le fait d'apparier la procédure ou non.

Les variables suivantes sont conservées dans l'ACM pour la police nationale :

- l'année de la procédure (10 modalités),
- le département du service ayant enregistré la procédure (104 modalités),
- l'indicatrice de présence d'au moins un mis en cause dans la procédure (2 modalités),
- le code de la direction de rattachement du service ayant enregistré la procédure (6 modalités),
- le type de saisine de la procédure (5 modalités).

Les variables suivantes sont conservées dans l'ACM pour la gendarmerie nationale :

- l'année de la procédure (6 modalités),
- le département du service ayant enregistré la procédure (104 modalités),
- l'indicatrice de présence d'au moins un mis en cause dans la procédure (2 modalités).

Pour la police nationale, les modalités suivantes sont très contributives des premiers axes factoriels : le fait que la procédure soit générée par la préfecture de Police de Paris (variable direction de rattachement du service), par un service du département 75 (Paris) ou 93 (Seine-Saint-Denis) pour le premier axe, la DTPN, Direction territoriale de la police nationale¹³ (variable direction de rattachement du service), ou par un service relevant des collectivités Outre-mer (COM) pour le second axe, la présence ou l'absence de mis en cause dans la procédure pour le troisième axe. Sur le premier axe, les procédures rattachées à la préfecture de Police de Paris, aux départements de services 75 et 93 sont négativement corrélées avec le fait que la procédure soit appariée. À l'inverse, sur le second axe, les procédures rattachées à la DTPN ou à un service des COM sont positivement corrélées avec le fait que la procédure soit appariée. Enfin, sur le troisième axe, les procédures comportant au moins un mis en cause sont positivement corrélées avec le fait que la procédure soit appariée.

Pour la gendarmerie nationale, les modalités suivantes sont très contributives des premiers axes factoriels : la présence ou l'absence de mis en cause dans la procédure pour le premier axe, l'année de procédure 2021 pour le second axe, l'année de procédure 2017 ou le département de service 17 (Charentes-Maritime) pour le troisième axe. Sur le premier axe, les procédures comportant au moins un mis en cause sont positivement corrélées avec le fait que la procédure soit appariée. À l'inverse, sur le second axe, les procédures créées en 2021 sont négativement corrélées avec le fait que la procédure soit appariée. Enfin, sur le troisième axe, les procédures créées en 2017 ou enregistrées dans le département 17 sont positivement corrélées avec le fait que la procédure soit appariée.

En conclusion, le non-appariement des procédures de violences conjugales paraît être caractérisé par une plus forte proportion de procédures sans mis en cause. Pour la police nationale, les procédures enregistrées par la préfecture de Police de Paris semblent moins bien appariées, alors que celles enregistrées dans les DROM semblent mieux appariées.

Analyse de la présence d'un mis en cause sur l'appariement

La **Figure 25** présente les taux d'appariement des procédures de violences conjugales 2022 pour la police nationale et la **Figure 26** des procédures de violence conjugales 2022 de la gendarmerie nationale, en fonction de la présence ou non d'un mis en cause dans la procédure.

• Figure 25. Statut d'appariement des procédures 2022 de police nationale selon la présence d'un mis en cause (en %)

	Présence de mis en cause	Pas de mis en cause	Total
Pas appariée	8,8 (6 668)	53,6 (25 121)	25,9 (31 789)
Appariée	91,2 (69 057)	46,4 (21 718)	74,1 (90 775)

Note de lecture : 8,8 % des procédures de la police nationale 2022 avec au moins un mis en cause ne sont pas appariées.

Source : bases Infractions, Victimes et Mis en cause 2022.

Dans le cas de la police nationale, les procédures de violences conjugales avec mis en cause sont appariées dans 91,2 % des cas, contre 46,4 % des cas quand la procédure n'a pas de mis en cause.

• Figure 26. Statut d'appariement des procédures 2022 de gendarmerie nationale selon la présence d'un mis en cause (en %)

	Présence de mis en cause	Pas de mis en cause	Total
Pas appariée	6,3 (3 680)	42,7 (8 499)	15,6 (12 179)
Appariée	93,7 (54 502)	57,3 (11 424)	84,4 (65 926)

Note de lecture : 6,3 % des procédures de la gendarmerie nationale avec au moins un mis en cause ne sont pas appariées.

Source : bases Infractions, Victimes et Mis en cause 2022.

Dans le cas de la gendarmerie nationale, les procédures de violences conjugales avec mis en cause sont appariées dans 93,7 % des cas, contre 57,3 % des cas quand la procédure n'a pas de mis en cause.

L'absence d'un mis en cause dans la procédure paraît donc jouer un rôle dans son appariement avec les données de la justice. Les procédures sans mis en cause sont mieux appariées en gendarmerie nationale qu'en police nationale. Il est possible que les procédures sans mis en cause soient plus souvent soumises à des transmissions de procédures à d'autres services, ou à des jonctions à des procédures existantes¹⁴. De ce fait, certaines procédures pourraient ne pas se retrouver dans le système d'information de la Justice. Il est également possible que l'absence de mis en cause rende complexe les poursuites judiciaires et que ces procédures soient directement classées sans suite, sans être enregistrées dans Cassiopée, ou encore que certaines procédures sont en attente pour être transmises à la Justice.

¹³ DTPN, administration déconcentrée de la police nationale présente dans certains territoires de la France d'outre-mer

¹⁴ Expertise qualitative menée sur un échantillon de procédures.

Analyse de l'impact de l'enregistrement par la
préfecture de Police de Paris ou par la DTPN
sur l'appariement

La **Figure 27** présente les taux d'appariement des procédures de violences conjugales 2022 pour la police nationale en selon qu'elles soient enregistrées par la préfecture de Police de Paris ou non.

Les procédures de violences conjugales enregistrées par la préfecture de Police de Paris sont appariées dans 60,5 % des cas, contre 77,0 % des cas quand la procédure est enregistrée hors préfecture de Police de Paris.

• **Figure 27. Statut d'appariement des procédures 2022 de police nationale selon l'enregistrement par la préfecture de Police de Paris (en %)**

	Enregistrement hors préfecture de Police de Paris	Enregistrement à la préfecture de Police de Paris	Total
Pas appariée	23,0 (23 080)	39,5 (8 709)	25,9 (31 789)
Appariée	77,0 (77 454)	60,5 (13 321)	74,1 (90 775)

Note de lecture : 23,0 % des procédures de la police nationale 2022 enregistrées hors de la préfecture de Police de Paris ne sont pas appariées.

Source : bases Infractions, Victimes et Mis en cause 2022.

La **Figure 28** présente les taux d'appariement des procédures de violences conjugales 2022 pour la police nationale en fonction du fait qu'elles soient enregistrées par la DTPN¹⁵ (Direction territoriale de la police nationale) ou non.

Les procédures de violences conjugales enregistrées par la DTPN sont appariées dans 81,5 % des cas, contre 74,0 % des cas quand la procédure est enregistrée par un autre service.

L'impact de l'enregistrement par la préfecture de Police de Paris ou par la DTPN sur le taux d'appariement total est à nuancer puisque le nombre de procédures par rapport à l'ensemble est assez faible.

• **Figure 28. Statut d'appariement des procédures 2022 de police nationale selon l'enregistrement à la DTPN (en %)**

	Enregistrement hors DTPN	Enregistrement à la DTPN	Total
Pas appariée	26,0 (31 547)	19,0 (242)	25,9 (31 789)
Appariée	74,0 (89 740)	81,5 (1 035)	74,1 (90 775)

Note de lecture : 26,0 % des procédures de la police nationale 2022 enregistrée hors DTPN ne sont pas appariées.

Source : base Infractions 2022.

Conclusion

L'enjeu méthodologique de l'appariement entre les données des SSM Sécurité Intérieure et Justice est de rapprocher des unités statistiques différentes utilisées dans les deux systèmes d'information : la procédure côté intérieur et l'affaire côté justice.

Pour ce faire, les procédures enregistrées par les FSI sont regroupées en paquets de procédures liées à une même situation de faits dénoncés. Ces procédures sont regroupées via un algorithme de théorie des graphes, utilisant les informations de la procédure d'origine ainsi que le CSNS des victimes et des mis en cause.

Il existe un identifiant commun aux deux systèmes d'information (le triplet), parfois manquant ou de mauvaise qualité : selon les années, entre 70,9 % et 75,5 % des procédures de violences conjugales enregistrées par les FSI sont appariées avec au moins une affaire Cassiopée. Au-delà du triplet, le CSNS des mis en cause est utilisé pour rapprocher des procédures et des affaires. Avec cette méthode d'appariement supplémentaire, selon les années, entre 74,9 % et 79,0 % des procédures de violences conjugales enregistrées par les FSI sont appariées avec au moins une affaire Cassiopée, soit une hausse comprise entre 3 et 4 points de pourcentage par rapport à l'utilisation seule du triplet. Enfin, le regroupement des procédures en paquets permet d'améliorer l'appariement, un paquet de procédures étant

considéré comme apparié si au moins une procédure de ce paquet est appariée. Ainsi, selon les années, entre 77,4 % et 82,1 % des paquets de procédures de violences conjugales sont appariés avec au moins une affaire Cassiopée, soit une hausse comprise entre 2 et 3 points de pourcentage par rapport à l'appariement des procédures seules.

En conséquence, environ une procédure sur cinq n'est pas appariée. Pour juger du caractère biaisé ou non de l'échantillon des procédures et affaires appariées, il faut déterminer s'il existe des caractéristiques spécifiques aux procédures non appariées par rapport aux procédures appariées sur le champ des violences conjugales. L'analyse des procédures non-appariées montre que les procédures qui ne comportent pas de mis en cause sont moins bien appariées, ainsi que les procédures enregistrées en région parisienne dans une moindre mesure.

Les travaux de caractérisation du non appariement et d'analyse de la qualité de l'appariement réalisé se poursuivront, via notamment une étude de cas sur des procédures non appariées, nécessaires pour faire la part entre les échecs d'appariement (i.e. des cas où une procédure est associée à une affaire mais le processus d'appariement n'a pas été en mesure de les identifier) et les situations où aucune affaire n'est en réalité associée aux procédures non appariées.

15 Administration déconcentrée de la police nationale présente dans certains territoires de la France d'outre-mer

Pour en savoir plus

- Brunin, L., Guedj, H., Le Rhun, B., 2019, Comparaison des statistiques Sécurité et Justice, le contentieux des violences conjugales, SSMSI, Interstats Méthode, n° 16.
- Bénichou, Y.-L., Espinasse, L., Gilles, S., 2023, Le code statistique non signifiant (CSNS) : un service pour faciliter les appariements de fichiers. Courrier des statistiques, n° 9.
- Groupe de travail SSMSI SDSE, 2016, Pour une méthodologie d'analyse comparée des statistiques Sécurité et Justice : l'exemple des infractions liées aux stupéfiants, SSMSI, Interstats Méthode, n° 8.
- Carrasco V., Lévêque E., 2025, Violences conjugales entre 2018 et 2023 : de la victime enregistrée à la décision de justice, SSER, Infostat Justice, n° 206, décembre.
- Briand A., Claude N., Koubi M., 2025, Méthodologie d'appariement entre les données administratives des forces de sécurité intérieure et celles de la justice, SSMSI, Interstats Méthode, n° 30.
- Makdessi Y., & Carrasco V., 2022, Rapprochement des données du SSMSI et de la SDSE, exemple des homicides. Présentation au Cnis, Commission Services publics et services aux publics.
- Salembier L., 2024a, La mesure du nombre d'homicides et de tentatives d'homicide enregistrés par la police et la gendarmerie depuis 2016, SSMSI, Interstats Méthode, n° 25.
- Salembier L., 2024b, Les bases statistiques du SSMSI sur la délinquance enregistrée, SSMSI, Interstats Méthode, n° 26.

Découvrez nos collections

- Infos Rapides Justice
- Infostat Justice
- Dossier Méthode
- Chiffres clés de la Justice
- Références Statistiques Justice
- Rapport d'études

[Site Internet du SSER](#)



Les données des tableaux, cartes et graphiques associés à cette étude, ainsi que des données sont disponibles sur Interstats, le site internet du SSMSI : www.justice.gouv.fr/documentation/etudes-statistiques