

# Données et documents administratifs : quels usages et implications pour la recherche en sciences sociales ?

### Appel à communications

Journée d'études organisée par le RT20 « Méthodes » de l'Association Française de Sociologie (AFS)

**Lundi 16 novembre 2026**

Centre des colloques, auditorium 150 - Campus Condorcet  
Place du Front Populaire, 93300 Aubervilliers

### Table des matières

Argumentaire .....	1
Modalités de soumission .....	6
Comité d'organisation.....	7
Références bibliographiques .....	7
Partenaires.....	8

### Argumentaire

Le recours des recherches en sciences sociales aux informations provenant de sources administratives n'est pas nouveau et prend plusieurs formes, que l'on songe à la contextualisation d'un phénomène en mobilisant des données statistiques ou des textes réglementaires ; ou au recours aux documents administratifs en tant qu'archives, pour enquêter sur des périodes et personnes auxquelles les chercheur-es n'ont pas directement accès. D'autres usages, relevant aussi bien de techniques quantitatives que qualitatives, prennent les dernières années de l'ampleur et méritent d'être interrogés, qu'il s'agisse de l'analyse secondaire des bases de données administratives ; de l'appariement de ces dernières aux données des enquêtes de recherche ; ou de l'utilisa-

tion des documents administratifs dans l'enquête ethnographique comme « outil d'analyse supplémentaire pour multiplier les points de vue sur [un] objet » (Béliard et Biland, 2008). Si l'usage des documents administratifs comme source est bien analysé par les historien·nes, cette journée propose de réfléchir aux usages des données et documents administratifs « vivants » dans les enquêtes en sciences sociales pour mieux cerner les enjeux, avantages, biais et contraintes induits par leur usage.

De quoi parle-t-on quand on évoque les données et documents administratifs ? Dans cet appel, les sources administratives désignent à la fois les enquêtes par questionnaire produites par la statistique publique pour « décrire la société » (Desrosières, 2005) comme l'Enquête annuelle de recensement, ou l'enquête Emploi, ainsi que les autres bases de données statistiques produites par des administrations et destinées informer les décideurs, les gestionnaires ou le public. Elles peuvent aussi désigner le produit des interactions entre les administrations et leur public : dossiers CAF, inscriptions à France Travail, demandes de titres de séjour, réponses à des formulaires en ligne, etc. D'autres sont produites par des professionnel·les des administrations qui documentent leurs activités soit pour informer les usager·es (dossiers scolaires, bulletins, etc.), soit pour les contrôler (rapports des inspecteurs de l'Éducation nationale, vérifications fiscales ou sociales, etc.), soit pour mieux organiser le suivi de leurs activités (dossiers de santé, etc.). Un certain nombre de ces données et documents administratifs peuvent être qualifiés de « vivants », c'est-à-dire qu'ils se définissent par leur caractère actuel, évolutif et opérationnel, étant encore activement utilisées ou produites par l'administration au moment où les chercheur·euses en sciences sociales s'en saisissent. À ce titre les conditions juridiques de l'accès à ces éléments ont été redéfinies : l'encadrement juridique entamé dans le XXe siècle — avec notamment la loi Informatique et libertés<sup>1</sup> puis le Règlement général sur la protection des données (RGPD) en 2018 — s'est également accompagnée d'une politique d'ouverture des données publiques<sup>2</sup> dans les années 2010, laquelle rend aussi nécessaire un travail d'ouverture de la part des chercheur·euses (Stérin et Noûs, 2019). Cela a ainsi donné lieu à des réflexions spécifiques sur les mesures à prendre pour faciliter l'accès des chercheur·euses aux données administratives (Cnis, 2017). Cette nouvelle offre administrative est par ailleurs concomitante du « moment *big data* des sciences sociales » (Bastin et Tubaro, 2018), voyant apparaître de nouvelles technologies informatiques permettant la massification et l'exploitation des nouveaux types de données issues des administrations. Toutefois, l'essor et la diversification des documents et données administratifs ne vont pas sans soulever des interrogations quant à leur mobilisation dans des travaux de recherche, ce type de matériau (à l'exception majeure des enquêtes de la statistique publique) étant le plus souvent élaboré à des fins autres que la recherche. Si ces données et documents offrent de nouvelles

---

<sup>1</sup> Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés individuelles.

<sup>2</sup> Voir la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique ainsi que celle, antérieure, sur la liberté d'accès aux documents administratifs (loi n° 78-753 du 17 juillet 1978).

pistes de recherche et une source de connaissance indéniable, leur périmètre, l'information contenue, leurs significations et leur forme ne sont pas toujours adaptés aux questionnements des chercheurs et chercheuses en sciences sociales.

Organisée par le Réseau thématique « Méthodes » (RT 20) de l'AFS, cette journée vise à confronter les approches méthodologiques dans l'usage des données et documents administratifs, s'intéressant aussi bien aux enquêtes quantitatives qu'aux enquêtes qualitatives. Elle fait appel à des travaux empiriques, rendant compte de la fabrique des sciences sociales, tout en invitant à la réflexivité sur ce que la recherche fait sur ces données et inversement. Pluridisciplinaire, elle s'adresse principalement aux chercheuses et chercheurs en sciences sociales, mais aussi aux statisticien·nes ou aux historien·nes travaillant sur l'histoire des données et documents administratifs et leurs relations avec la recherche.

Les propositions de communication pourraient ainsi combiner trois axes de réflexion :

### **Axe 1 : Se réappropriier un matériau parfois hétérogène et non structuré à des fins de recherche**

L'appropriation par les chercheur·euses des données et documents administratifs soulève un ensemble de questions liées à la nature même de matériaux produits pour d'autres finalités que la recherche. Ces données sont d'une part générées par une pluralité d'administrations, et en leur sein, une pluralité d'acteurs et d'actrices. Pensées pour le suivi de populations, la gestion quotidienne, le contrôle, les études ou les justifications administratives, voire commerciales, ces données ne sont pas toujours compatibles avec les exigences des sciences sociales. Comment les chercheur·euses travaillent-ils et elles avec un tel matériau ? En outre, les « administré·es » ne sont pas des réceptacles passifs, et peuvent modifier leur attitude ou réponses selon la représentation qu'ils et elles se font d'une administration et de leurs intérêts. Plus généralement, leurs réponses peuvent traduire des manières de penser qui diffèrent des catégories envisagées par les administrations (Bessière et Houseaux, 1997). Les données parviennent ainsi aux chercheurs et chercheuses dans un état qui reflète aussi bien les logiques institutionnelles ou organisationnelles sous-jacentes que les stratégies des administré·es, et nécessitent un travail d'appropriation, plus ou moins élaboré, pour être analysées. Les communications pourront revenir sur les activités de traitement de ces données mises en place par les chercheur·euses pour s'appropriier les données.

Face à ces matériaux hétérogènes, souvent volumineux, parfois lacunaires ou redondants, ou tout simplement conçus à partir d'autres préoccupations, questions et formes d'administration de la preuve, les chercheur·euses doivent déployer tout un ensemble de pratiques d'appropriation visant à transformer ces données en ressources analytiques compatibles avec leurs recherches. Qu'il s'agisse des résultats d'une enquête par questionnaire, d'un dossier individuel comportant plusieurs années de correspondances, d'une base de données issue de formulaires remplis par les administré·es ou enrichie automatiquement par différents services, ou d'un fi-

chier d'usage générant des milliers d'événements numériques, les chercheur·euses doit déterminer ce qui mérite d'être retenu, reconstitué ou, au contraire, mis à distance. Dans cette perspective, les pratiques de sélection des données pertinentes, de recodage, de tri, de nettoyage, d'extraction ou d'annotation deviennent des opérations proprement scientifiques : elles participent de la construction des matériaux et orientent les résultats possibles. Les propositions de communication pourront éclairer les opérations concrètes par lesquels les chercheur·euses préparent, transforment et analysent ces matériaux, et mettre en évidence les effets de ces choix sur la construction du corpus et les analyses ensuite produites. Nous espérons ainsi, en analysant de manière critique ces usages, mieux saisir les objectifs de ce recours aux sources administratives, leurs avantages mais aussi leurs particularités et leurs risques au regard des problématiques et des méthodes de recherche qu'elles impliquent.

Cette journée sera ainsi l'occasion de mettre la focale sur les institutions productrices de données, pour tenir compte des multiples cadres de production, des objectifs, des dispositifs de collecte et de conservation des données. Quels rapports les producteurs entretiennent aux données et avec les chercheurs et les chercheuses ? Qu'est-ce que les conditions de production des données font aux méthodes et aux résultats de recherche ? Si la journée se concentre majoritairement sur les productions administratives du secteur public, des communications sur d'autres environnements organisationnels, comme les entreprises ou les associations, seront les bienvenues. Notre appel s'adresse également aux travaux de recherche en la matière dans d'autres pays. Ceci permettrait en outre de mettre en perspective les expériences françaises au regard des processus de normalisation internationaux qu'il s'agisse de nouveaux types de données comme les *big data* (Bastin et Tubaro, 2018), de mesures juridiques comme la RGPD ou de processus de standardisation (Institut de statistique de l'Unesco, 2024).

## **Axe 2 : Ce que les données et documents administratifs disent du travail des administrations**

Si la façon dont les données ont été construites vient avec son lot d'inconvénients pour les recherches en sciences sociales, cela peut donner à voir la façon dont l'administration conçoit son travail et les individus qu'elle administre. Tout d'abord, les communications pourront se pencher sur les situations de travail et les motivations qui animent les personnes en charge de la production des données administratives. Les conditions de leur production ont un effet sur la qualité des données (Becker, 2020) comme le donnent à voir les travaux sur les employées au guichet des services publics dans les quartiers populaires (Siblot, 2006). Les communications pourront explorer le lien entre les contraintes professionnelles des producteurs de données et la nature et le type d'information consignées. Par exemple, les greffier·es chargé·es de la collecte d'informations auprès des détenu·es pour l'élaboration des statistiques pénitentiaires font face à des contraintes temporelles importantes, doublées d'une absence de formation à la collecte de données (Henneguelle, 2017 ; Denis et Goëta, 2017).

Parallèlement, il faut tenir compte des propriétés socioprofessionnelles des personnes qui produisent les données et de la division du travail entre elles. Quel est l'effet de la collecte, de la

conception au codage, sur les données ? Les données et documents administratifs sont le fruit de multiples opérations de classement par les administrations elles-mêmes. L'étude des données et documents administratifs permet ainsi de mettre en évidence les formes de catégorisation à l'œuvre au sein des institutions, et qui participent de la soumission des individus à un ordre institutionnel et politique. C'est par exemple que qu'illustrent les documents produits par des conseillères en économie sociale et familiale, qui traduisent des normes comptables administratives en décalage avec les pratiques des familles pauvres qu'elles accompagnent (Perrin-Heredia, 2011). De même, l'analyse des catégories utilisées par l'Insee montre comment, lors du rapatriement depuis l'Algérie, l'État français a « continué de séparer les populations originaires d'Algérie selon un critère ethno-religieux alors que cette distinction n'avait plus d'existence légale » (Couto, 2023). Les communications pourront ainsi porter sur ce travail de catégorisation des populations et sa confrontation avec les catégories des administré·es. D'ailleurs, le travail de catégorisation n'est pas forcément univoque. Selon Béliard et Biland (2008), l'usage de ce type de matériaux permet aussi de révéler les rapports de pouvoir entre les acteurs impliqués dans le traitement des individus, et « la manière dont l'institution prend les décisions et parvient à parler d'une seule voix (...) en faisant taire les dissonances ». L'étude du matériau ici analysé pourrait enfin révéler la manière dont les normes et les catégorisations nationales sont traduites au niveau administratif local (Béliard et Biland, 2008). Nous encourageons ainsi les communications à documenter autant les catégorisations administratives et leur production que ce qu'elles produisent sur les administré·es y compris en matière de soumission à un ordre politique.

### **Axe 3 : Enjeux éthiques et pratiques**

Parce que leur production implique des rapports de pouvoir et des contraintes institutionnelles, les données et documents administratifs soulèvent, au-delà de leur analyse, des enjeux éthiques et pratiques centraux pour la recherche. Dans cette perspective, il importe de noter que ces données et documents sont, le plus souvent, à caractère personnel et peuvent parfois être récoltées sans l'assentiment des personnes concernées. Plusieurs questions éthiques se posent donc quant à leur collecte, leur conservation et leur ré-exploitation dans le cadre de la recherche. Si certaines sont circonscrites par des cadres réglementaires<sup>3</sup>, d'autres sont du ressort des chercheurs ou des chercheuses, en fonction des objectifs, des contours et des situations prises en compte par une recherche donnée. Dans un souci de protection des personnes, le respect de l'anonymat des individus dans les publications est un enjeu majeur auquel s'ajoute celui de ne pas collecter des informations qui ne seraient pas strictement nécessaire à la recherche, comme le souligne le principe de minimisation de la CNIL. De ce point de vue, les possibilités d'interconnexion ou d'appariement de sources administratives restent un point délicat dont les implications méritent d'être davantage interrogées. Les controverses sur la création d'un identifiant unique des personnes physiques sont connues. Mis en avant pendant l'Occupation, ce numéro, porteur de la mémoire sinistre du gouvernement de Vichy, a été rejeté à la Libération, puis dans les années

---

<sup>3</sup> Il s'agit notamment de la loi n°51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques et de la loi de 2018 relative à la protection des données personnelles.

1970 où il a été réenvisagé (Poulain, 2022). Ces dernières années, la volonté de promouvoir l'appariement de données administratives a donné lieu à la création du Code statistique non signifiant (CSNS), numéro qui n'apporte aucune information sur la personne à laquelle il est rattaché (Bénichou et al., 2023). Dans ce contexte, les communications proposées, pourront apporter un éclairage réflexif sur le sens, l'impact et les possibilités des pratiques en matière de mobilisation, d'appariement et de protection des données personnelles dès lors que sont mobilisées des données administratives pour la recherche en sciences sociales.

De la même façon, la réutilisation des données personnelles au-delà de la finalité pour laquelle elles ont été collectées peut être questionnée. Dans quelle mesure et sous quelles conditions sommes-nous autorisé·es à les transporter au-delà du contexte et de la finalité de leur collecte ? Même si, dans certains cas, les personnes peuvent être initialement informées d'une possible réutilisation pour des fins de recherche scientifique, les termes et les objectifs de cette recherche leur échappent. L'information sur cette possible réutilisation n'est pas toujours perceptible par les répondant·es, peu averti·es des procédures de circulation des données. Les communications se centrant sur ces questionnements sont vivement encouragées.

De surcroît, le caractère personnel du matériau ici considéré peut rendre l'accès plus difficile, coûteux et chronophage. Les négociations avec les administrations peuvent être longues et l'accord risque d'être associé à une réorientation de la recherche. L'accès aux données statistiques s'effectue de plus en plus souvent via le Centre d'accès sécurisé aux données (CASD), ce qui génère des inégalités entre chercheur·es selon leur statut et leur rattachement. L'accès au matériau peut être assez restrictif, aussi bien par son coût financier que pour s'assurer du respect de la confidentialité, ou pour ne pas déranger le travail quotidien de l'administration. C'est par exemple le cas de certaines recherches sur les dossiers scolaires ou médicaux qui impliquent que les chercheur·es consultent les documents dans un lieu défini (se rapprochant ainsi des travaux historiques sur archives) et potentiellement sous le regard de l'administration étudiée (Cayouette-Remblière, 2016 ; Mathieu, Froger-Lefebvre et Dall'Agnola, 2025). Les communications pourront ainsi se pencher sur les contraintes pratiques d'accès aux données et documents administratifs et éclairer les conditions de possibilité de ce type de recherche (notamment au regard du positionnement institutionnel du ou de la chercheur·euse).

## Modalités de soumission

Les propositions de communication, d'une taille maximale de 3 500 signes (espaces compris, hors bibliographie), sont à déposer **le lundi 4 mai au plus tard** sur le site suivant : <https://donnees-admin.sciencesconf.org>. Elles doivent inclure, outre le titre et le résumé, vos nom(s), rattachement(s) institutionnel(s) et adresse(s) électronique(s).

Les communications dureront environ une vingtaine de minutes.

Les propositions de communication seront évaluées par les membres du bureau du RT20. Les résultats seront transmis aux auteur·ices mi-juillet.

## Comité d'organisation

- Paul COCHET (Ined, SES),
- Julia DESCAMPS (Ined, UR8),
- Efi MARKOU (Ined, SES),
- Lucas PAGE PEREIRA (ENS Paris-Saclay, IDHE.S),
- Vic SESSEGO (Université de Versailles-Saint-Quentin-en-Yvelines, PRINTEMPS),
- Marion VALARCHER (Université de Tours, CITERES)

Pour toute information vous pouvez nous contacter à l'adresse :

[donnees-admin@sciencesconf.org](mailto:donnees-admin@sciencesconf.org)

## Références bibliographiques

Bastin, Gilles et Paola Tubaro. « Le moment big data des sciences sociales », *Revue française de sociologie*. 2018, vol.59 no 3. p. 375-394.

Becker, Howard S. « Données administratives » *Faire preuve*. Paris : La Découverte. 2020, p. 137-174. (Repères).

Béliard, Aude et Émilie Biland. « Enquêter à partir de dossiers personnels. Une ethnographie des relations entre institutions et individus », *Genèses*. 2008, vol.70 no 1. p. 106-119.

Bessière, Céline, et Frédérique Houseaux. « Suivre des enquêteurs ». *Genèses. Sciences sociales et histoire*. 1997, vol. 29, n° 1, p. 100-14.

Cayouette-Remblière, Joanie. *L'école qui classe. 530 élèves du primaire au bac*. Paris : Presses Universitaires de France. 2016. 312 p. (Le Lien social).

Conseil national de l'information statistique (CNIS). *L'accès des chercheurs aux données administratives. État des lieux et propositions d'actions*. Rapport 147. Conseil national de l'information statistique. 2017.

Couto, Marie-Paule. « Naturels » ou « naturalisés » ? Le travail d'identification des rapatriés d'Algérie par l'État et ses effets sur la recherche en sciences sociales », *Sociologie*. 2023, vol.14 no 1. p. 49-70.

Denis, Jérôme et Goëta, Samuel. « La fabrique des données brutes : Le travail en coulisses de l'open data ». In : Mabi, Clément, Plantin, Jean-Christophe et Monnoyer-Smith, Laurence (éd.), *Ouvrir, partager, réutiliser : Regards critiques sur les données numériques*. Paris : Éditions de la Maison des sciences de l'homme. faber. 2017, ISBN 978-2-7351-2386-5.

Desrosières, Alain. « Décrire l'État ou explorer la société : les deux sources de la statistique publique », *Genèses*. 2005, vol.58 no 1. p. 4-27.

Henneguelle, Anaïs. « La fabrique des statistiques pénitentiaires », *Criminocorpus. Revue d'histoire de la justice, des crimes et des peines*. 2017, no 8.

Institut de statistique de l'UNESCO (UIS). « Données administratives : défis et solutions proposées ».

Mathieu, Marie, Juliette Froger-Lefebvre, et Anne-Lise Dall'Agnola. « Des conditions sociales d'accès et de traitement de dossiers médicaux à leurs usages sociologiques. Retour sur une incursion collective au sein de services d'oncogénétique. » Communication au RT20 lors du congrès de l'Association Française de Sociologie à Toulouse. 2025.

Poulain, Claude. « Le projet SAFARI (1970-1974) », *Terminal. Technologie de l'information, culture & société*. 2022, no 134-135.

Siblot, Yasmine. *Faire valoir ses droits au quotidien. Les services publics dans les quartiers populaires*. Paris : Presses de Sciences Po. 2006. 476 p.

Stérim, Anne-Laure, et Camille Noûs. « Ouverture des données de la recherche : les mutations juridiques récentes ». *Tracés. Revue de Sciences humaines*. 2019, #19, p. 37-50.

## Partenaires



« Ce travail a bénéficié d'une aide de l'État au titre de France 2030 portant la référence "ANR-11-IDEX-003". »