

RENDRE COMPTE DU TRAVAIL D'INSTAURATION ET D'ASSEMBLAGE DES DONNÉES EN SANTÉ : LE CAS DE REGISTRES DES CANCERS EN FRANCE

[Anja Martin-Scholz](#), [Anne Mayère](#), [François Lambotte](#)

Presses universitaires de Bordeaux | « [Communication & Organisation](#) »

2021/1 n° 59 | pages 215 à 230

ISSN 1168-5549

ISBN 9791030006957

DOI 10.4000/communicationorganisation.10195

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-communication-et-organisation-2021-1-page-215.htm>

Distribution électronique Cairn.info pour Presses universitaires de Bordeaux.

© Presses universitaires de Bordeaux. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.



Communication et organisation

Revue scientifique francophone en Communication organisationnelle

59 | 2021

La communication constitutive des organisations :
émergence et innovations

Rendre compte du travail d'instauration et d'assemblage des données en santé : le cas de Registres des cancers en France

Making the work of establishing and assembling health data visible : the case of French cancer registries

Anja Martin-Scholz, Anne Mayère et François Lambotte



Édition électronique

URL : <https://journals.openedition.org/communicationorganisation/10195>

DOI : 10.4000/communicationorganisation.10195

ISSN : 1775-3546

Éditeur

Presses universitaires de Bordeaux

Édition imprimée

Date de publication : 1 juin 2021

Pagination : 215-230

ISSN : 1168-5549

Distribution électronique Cairn



CHERCHER, REPÉRER, AVANCER.

Référence électronique

Anja Martin-Scholz, Anne Mayère et François Lambotte, « Rendre compte du travail d'instauration et d'assemblage des données en santé : le cas de Registres des cancers en France », *Communication et organisation* [En ligne], 59 | 2021, mis en ligne le 01 janvier 2025, consulté le 28 juillet 2021. URL :

<http://journals.openedition.org/communicationorganisation/10195> ; DOI : <https://doi.org/10.4000/communicationorganisation.10195>

© Presses universitaires de Bordeaux

Rendre compte du travail d'instauration et d'assemblage des données en santé : le cas de Registres des cancers en France

Making the work of establishing and assembling health data visible: the case of French cancer registries

Anja Martin-Scholz, Anne Mayère, François Lambotte

Résumé : La montée en puissance des discours promouvant le « Big Data » laisse à penser qu'il suffirait d'accumuler et de croiser un nombre suffisant de « données » pour avoir une vue exhaustive des phénomènes. En prenant appui sur une recherche en cours sur le terrain d'organisations en charge de la production de Registres des cancers, nous proposons de spécifier le processus d'instauration des données en articulation avec leurs assemblages. Il s'agit ainsi de caractériser ce qui fonde les modes d'existence des données et de leurs assemblages hétérogènes et instables. Dépassant la focale sur les finalités affichées, il s'agit de mieux comprendre ce qui est à l'œuvre dans la fabrique des données, à travers des zooms avant et arrière. Sont ainsi identifiés des enjeux en termes de savoir mais aussi en termes politiques, économiques et sociaux.

Mots-clés : santé, données, bases de données, autorité, instauration, assemblage.

Abstract: The rise of discourses promoting “Big Data” leads one to believe that it would be sufficient to accumulate and cross-reference a sufficient number of “data” to have an exhaustive view of a phenomenon. Based on ongoing research in the field of organizations in charge of producing cancer registries, we propose to specify the process of data creation in articulation with their assemblage. The aim is to characterise what underlies the modes of existence of data and their heterogeneous and unstable assemblages. Going beyond the focus on the stated purposes, the aim is to better understand what is at work in the manufacture of data, by zooming in and out. The issues at stake are thus identified in terms of knowledge, but also in political, economic and social terms. Translated with www.DeepL.com/Translator (free version).

Keywords: health, data, database, authority, elaboration, assembly.

Évaluer et avoir une idée précise du nombre de malades constitue un enjeu de santé publique. La question de ce qui est mesuré renvoie à l'importance d'un accord sur les définitions et les catégories (Desrosières 2008), en particulier dès qu'il s'agit de comparaison entre différentes régions ou avec d'autres pays. La période actuelle caractérisée par la pandémie du Covid-19 est plus qu'emblématique en la matière. Le manque d'homogénéité des définitions et d'exhaustivité du comptage à partir de ces définitions soulève le problème d'une impossible comparaison entre pays, quant aux chiffres de mortalité et quant aux courbes qui suivent l'évolution de la maladie.

La création, en 2019, du Health Data Hub ou « plate-forme d'exploitation des données de santé françaises » ayant « pour objectif de favoriser (leur) utilisation et de multiplier les possibilités (de leur) exploitation » (drees.solidarites-sante.gouv.fr) a confirmé la montée en puissance des discours et projets promouvant le « Big Data » en France. Le manifeste qui affiche son objectif fait entrevoir un horizon selon lequel l'accumulation de données pourrait permettre de donner une vue d'ensemble des phénomènes, à condition que ces données soient très nombreuses.

Analyses d'importants jeux de données en mobilisant les approches dites d'« intelligence artificielle » peut alors éclairer sur la stratégie thérapeutique à adopter, patient par patient [...]. À cette fin, il devient indispensable de constituer de grands jeux de données mobilisant des dossiers patients de plusieurs centres hospitaliers pour avoir une masse critique permettant de réaliser des inférences performantes et précises. De plus, dans un contexte de médecine ambulatoire, la recherche ne peut plus reposer uniquement sur les données des dossiers médicaux hospitaliers, elle doit également mobiliser ceux de la médecine de ville, ainsi que les données produites par les patients eux-mêmes. Par conséquent, elle sera de plus en plus difficile à réaliser sans recourir à une part d'automatisation pour extraire de l'information pertinente dans de grands jeux de données hétérogènes. (Health Data Hub, Mission de préfiguration, 12/10/2018, p. 9-10 publié sur solidarites-sante.gouv.fr)

Ces discours promeuvent de façon liée « l'intelligence artificielle » comme susceptible d'avancées sur des enjeux multiples, tant en termes de politiques publiques que de marchés (Lehmans 2017 ; Ménard & Mondoux 2018). Ce faisant, ils minimisent et invisibilisent le travail complexe des données requis en amont tant pour leur construction que pour leur intelligibilité (Denis & Goëta 2017). En effet, les constructions de données mettent en avant certaines catégories tout en en taisant d'autres (Suchman 1993). De plus, ces approches décontextualisent la qualité des données qui est le produit d'organisations, de processus organisants spécifiques (Mayère 2018), avec leurs logiques, en référence à des normes, selon des acceptations situées qui fondent leur autorité et leur auteurité (Wathelet 2015). Pour ouvrir la boîte noire et mettre en évidence les processus organisants à l'œuvre dans cette inter-organisation en constant chantier, nous proposons de mettre en évidence comment ces données sont issues d'interactions et de collaborations dans des réseaux denses et diversifiés d'acteurs et plus largement d'actants, dans des configurations inter-organisationnelles plurielles (Martin-Scholz 2019). Pour ce faire, nous prenons appui sur deux études de cas de Registres des cancers. Nous allons dans un premier temps

présenter le terrain et la méthode d'investigation. Dans un second temps, à partir de l'analyse du matériau ainsi rassemblé, nous allons caractériser plus avant ce qu'il en est du processus d'instauration des données et de leur assemblage. Il s'agit ainsi de documenter ce qui fonde les modes d'existence des données et de leurs assemblages hétérogènes et instables.

Terrain et Méthode : étudier la fabrique de données de « Registres des Cancers »

Cette publication s'appuie sur une recherche en cours sur les Registres des cancers de l'ouest, du sud-ouest et du sud de la France. Les Registres des Cancers sont des organisations ayant pour mission de réaliser « un recueil continu et exhaustif de données nominatives intéressant un ou plusieurs événements de santé dans une population géographiquement définie, à des fins de recherche et de santé publique, par une équipe ayant les compétences appropriées. » (arrêté du 6/11/1995 relatif au Comité National des Registres). Selon cet arrêté, les organisations en charge des Registres ont une double mission : d'une part, elles assurent une veille sanitaire concernant les cancers, en relevant l'incidence et la mortalité liées des cancers sur un territoire donné ; d'autre part, elles effectuent des études et des recherches à partir de leurs données, afin, notamment, d'évaluer l'organisation des parcours de soins, les campagnes de prévention ou encore l'influence des inégalités sociales sur l'incidence et/ou la survie. Elles sont ainsi évaluées sur la qualité des données, mais aussi et surtout sur les publications effectuées en leur sein.

Les trois organisations enquêtées sont des Registres généraux ayant une portée départementale ; sont enregistrés tous les diagnostics de cancer concernant des individus vivant dans le département en question, renseignés selon un socle de critères établi par des standards internationaux. Des informations complémentaires peuvent être rassemblées en fonction des programmes et des thématiques de recherche auxquels contribuent ces organisations.

Dans la perspective des Registres, un critère-clé de qualité tient à ce que toutes les occurrences des cas de cancer dans le département soient recensées de façon exhaustive ; il importe également de vérifier que les cas enregistrés soient bien des cancers selon la définition de l'OMS. Pour ce faire, les agents des Registres ont recours à différentes sources de notification et les croisent. En effet, aucune source n'est exhaustive ; tous les cancers ne font pas obligatoirement l'objet d'une biopsie, ni d'une hospitalisation.

Selon les acquis de l'ethnométhodologie (Coulon 2014), nous prenons acte de ce que les acteurs sont à même de formaliser une fabrique, d'en révéler la complexité et ainsi de donner des clés de lecture pour permettre aux chercheurs d'être équipés. Afin

de saisir le contexte de production des données relatives au cancer, nous avons d'abord réalisé une série d'entretiens semi-directifs approfondis auprès d'épidémiologistes en santé publique, de mathématiciens spécialisés dans le traitement de données en santé, d'un « data scientist » d'un CHU et d'un responsable d'un centre hospitalier spécialisé en oncologie. L'investigation s'est ensuite centrée sur les Registres de trois départements pour lesquels nous avons réalisé des entretiens auprès d'épidémiologistes de ces registres (cinq entretiens d'une durée totale de 7 heures). Les entretiens ont suivi une trame explorant les actions, les relations, et ainsi les conditions d'existence de la fabrique de données et de leurs assemblages. Par ailleurs, un ensemble de documents ont été rassemblés, tels que le manuel servant en appui au travail de codification ou les fiches de collecte d'informations. L'analyse croisée des entretiens et leur confrontation avec les documents mettent en lumière différents actants mobilisés dans ce travail. Il est prévu de poursuivre ces investigations par des « filatures » (*ibid.*) d'enquêtrices¹, en charge d'identifier, de trier et d'annoter les informations nécessaires à l'enregistrement de cas de cancer, et de secrétaires, en charge du codage des cas dans les bases de données, des Registres dans trois départements.

Une fabrique faite d'instauration et d'assemblage

Pour étudier plus en avant la fabrique des données, nous proposons de faire un double mouvement de « zoom avant/zoom arrière » (Nicolini 2009 ; Grosjean 2013). En effet, les activités situées au sein des organisations des Registres sont connectées à un ensemble d'éléments situés dans un « ailleurs » (Latour 2007). Il s'agit dès lors de prêter attention aux activités et aux pratiques locales pour comprendre comment les acteurs font sens en situation, mais aussi d'élargir la vision pour montrer comment ces activités sont connectées dans un réseau de pratiques et de formes organisationnelles (Grosjean 2013). À travers un mouvement de « zoom avant », nous allons explorer et décrire les différentes « opérations » (Denis & Goëta 2017) qui organisent le travail des enquêtrices. Denis et Goëta, dans leur article portant sur la fabrique des données ouvertes, identifient « une série d'opérations qui instaurent progressivement des informations aux contours flous en données » (*ibid.* : 4). Le mouvement de « zoom arrière » révèle, quant à lui, comment ces activités sont connectées à d'autres lieux et mobilisent des ressources organisationnelles autres. En étudiant les fabriques d'assemblage de données, nous mettons en évidence un réseau d'actants pluriels au rythme des opérations d'instauration et d'assemblage, dont nous allons explorer les dimensions et les articulations.

1- Du fait, que ces personnes soient, dans une très grande majorité, des femmes et dans un souci de lisibilité, nous avons décidé de maintenir la dénomination utilisée par les personnes rencontrées, à savoir « enquêtrice ».

L'instauration des données

Nous analysons l'*instauration*, en nous fondant sur la proposition de Denis (2018) et de Lambotte (2017), qui s'appuient eux-mêmes sur Latour (2010 ; 2014). Retravaillant cette notion à partir de Souriau, Latour montre comment cette notion permet de « saisir 'l'œuvre à faire' », sans devoir choisir entre celui qui fait et ce qui est fait, et en considérant ce en quoi « elle va vers l'existence » (*ibid.* : 18). Il se démarque ainsi de la conception constructiviste qui veut que « les faits soient construits ». Le savant ne construit pas les faits à partir de rien. Mais nous dit Latour, « s'ils sont instaurés par le savant ou par l'artiste, alors les faits comme les œuvres tiennent, résistent, obligent – et les humains, leurs auteurs, doivent *se dévouer* pour eux, ce qui ne veut pourtant pas dire qu'ils leur servent de simple conduit » (*ibid.* : 19). Ainsi dans l'instauration, il y a un risque pris, une existence croissante ; la focale est sur la modalité du faire exister, et non sur le seul résultat. Par ailleurs, la focale ne porte pas sur les acteurs, mais sur les actions (Lambotte 2017) et sur les relations, qu'il s'agit « d'enquêter de façon sérieuse » (Latour 2010 : 17).

Les données ne sont ni déjà-là, ni construites ex-nihilo, mais instaurées. L'idée selon laquelle les données attendraient patiemment sur les étagères d'être identifiées pour être « ouvertes » relève du fantasme (Denis 2018). Si les « données » des Registres ne sont pas déjà constituées en tant que telles, il faut donc que les agents des Registres les instaurent. Ils le font en construisant et reconstruisant des collaborations avec des interlocuteurs de différentes organisations, en explorant différents documents et fichiers, en construisant des continuités entre des indices pour identifier un « cas », en regroupant des cas toujours singuliers en autant de catégories établies par des définitions internationales.

Il peut co-exister différents registres d'instauration ; ainsi les professionnelles distinguent les données dites « de fonctionnement », qui permettent de tracer le cheminement des enquêtes, de celles qui formalisent les indices et se prêtent à catégorisations et agrégations.

L'instauration des données suppose d'identifier des traces, de croiser des sources, de transformer en données par la codification (Fig. 1). Elle est, de façon étroitement liée, affaire d'assemblage (Puig de la Bellacasa 2011). En effet, les données sont produites pour être assemblées, pliées et dépliées en différentes configurations, pour construire différentes intelligibilités. En cela, instauration et assemblage sont deux dimensions articulées du mode d'existence des données.

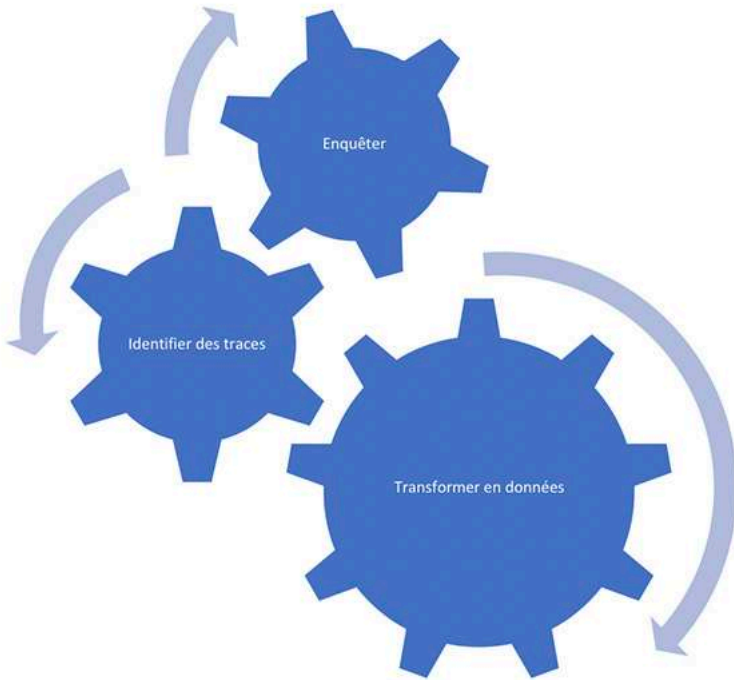


Figure 1: *instauration des données.*

Identifier des traces

Afin d'enregistrer les cas de cancers, les enquêtrices des Registres vont solliciter différents organismes, des « sources de notification », afin de recueillir auprès d'eux différents fichiers et documents qui leur permettront d'identifier des indices de cas « potentiels » de cancers. Des « sources » sont documentées, des liens sont établis avec des interlocuteurs dans les établissements de santé sur le territoire pour obtenir à échéance régulière des enregistrements sur les actes susceptibles de signaler de nouveaux cas. Des formats de fichiers et d'échange sont expérimentés, routinisés, reconstitués au gré des transformations. Comme le figure le schéma résultant de l'analyse de notre corpus (Fig. 2) et à la différence de Denis qui investigate l'instauration des données issues d'une source unique (Denis 2018), l'élaboration d'un registre repose sur de multiples sources de notification, élaborées par une diversité d'organisations. Avant d'être des données pour la veille sanitaire sur le cancer et/ou la recherche épidémiologique sur cet ensemble de maladies, les éléments, traces

et indices mobilisés peuvent être de différents ordres : ils peuvent être extraits de comptes rendus d'analyse de cellules, de dossiers constitués pour des réunions de concertation pluridisciplinaire, tels les DCC (dossiers communicants des cancers) ou encore des informations médico-économiques telles que celles rassemblées dans le PMSI (Programme de Médicalisation des Systèmes d'Information).

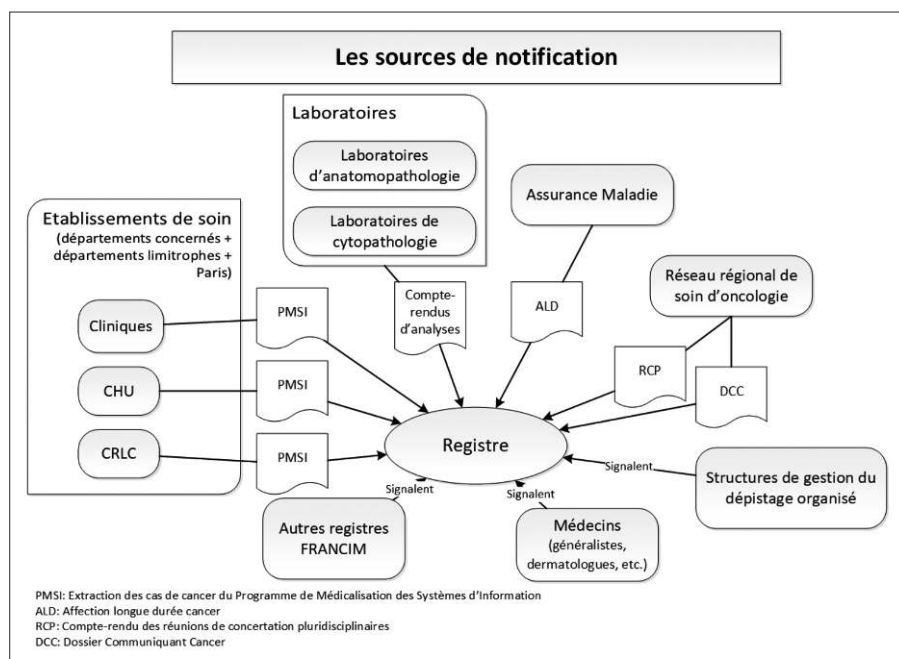


Figure 2: cartographie de la diversité des sources de notification et des organisations en interaction dans la fabrication des données.

Ainsi l'accès aux enregistrements, aux traces, prend forme par des médiations plurielles, des mises en relation négociées entre acteurs locaux et globaux, rendant possible une instauration progressive, participant à l'existence croissante des données sur les « cas » repérés.

Cette instauration repose sur une mise en relations qui articule de façon itérative des liens tissés et retissés sur un territoire donné, et des autorisations de mises en lien requérant l'intervention d'instances nationales voire internationales. Les documents en question ont trait à des données de santé ; pouvoir y accéder fait à la fois l'objet d'aller-retours entre les enquêtrices, de négociations entre les enquêtrices et les personnes en charge dans diverses organisations de leur enregistrement, mais aussi d'échanges à un niveau plus global, afin d'obtenir les autorisations requises.

Mener l'enquête

Une fois identifiés des patients, les enquêtrices vont mener des enquêtes afin (1) de vérifier l'identité des personnes et (2) de s'assurer qu'il s'agit bien d'un « nouveau cancer » ; les métastases et les récidives étant considérées par l'OMS comme un « prolongement » du cancer initial et donc non comptés comme « nouveau cancer ». Ce faisant, ici s'inscrit l'optique choisie pour recenser les cancers au sein des Registres ; ceci *structure* la fabrique des données, son périmètre premier. Les données ainsi instaurées n'incluent pas la caractérisation des soins effectués, les proportions de récidives pour chacun des cancers, ou la capacité de tel ou tel cancer à susciter des métastases dans tel ou tel organe. Les Registres ont pour mission de contribuer à la veille sanitaire, il s'agit dès lors d'identifier en priorité les facteurs de risques de cancer et les variations « anormales » du nombre de cas et ainsi contribuer aux politiques de santé publique. Dans cette perspective, le caractère systématique du recensement des cas est essentiel, ainsi que le respect des règles de codification, dans la mesure où ils conditionnent la fiabilité des mesures et leur comparabilité.

Le long et méticuleux travail des enquêtrices consiste ainsi à vérifier qu'il s'agit bien d'un nouveau cas, et de le documenter dans une nouvelle fiche, ou de rattacher de nouveaux éléments à une fiche préexistante et de la compléter. Ainsi, elles détectent et assemblent sous un nouveau format, la fiche d'investigation, ce qui était organisé selon d'autres finalités et d'autres formats, distribué entre différents acteurs et organisations, tels les laboratoires d'anatomopathologie, les hôpitaux ou les cliniques.

Ce travail d'enquête repose sur un chaînage précis constitué en autant d'indices susceptibles d'instaurer la validité et la caractérisation de chaque cas, et de son classement adéquat, avec pour exigence le caractère systématique de l'investigation, pour tracer également les « cas non standards », ce qui peut relever d'errances, ou d'aiguillages inattendus :

Responsable Registre 1 : Un cas standard, ça [ne] prend pas énormément de temps. Le problème, ce sont les gens qui sont arrivés à un endroit, ç'a pas marché, ça, ça prend énormément de temps.

Chercheur-e : Et ça se passe comment pour ces cas-là ?

Resp. 1 : Alors nous, ça part et ça revient. Quand les enquêtrices ramènent les cas, elles les donnent à la personne qui code. [...]. Et donc la personne qui code, elle va dire « je n'ai pas assez pour coder ». L'enquêtrice ne le sait pas nécessairement qu'il n'y a pas assez pour coder puisqu'elle, elle va collecter toute l'information qu'elle trouve dans l'endroit où elle est. Et elle ne sait pas si ailleurs, on n'aura pas d'autres infos qui seront suffisantes pour compléter. Donc elle arrive avec tout ce qu'elle a trouvé et elle le donne. Mais si ailleurs, c'était les mêmes informations et puis qu'il y a un trou, la personne qui code va dire « je n'ai pas assez d'informations, il y a un petit problème, je n'arrive pas à comprendre si c'était quelqu'un qui était métastatique, je ne suis pas sûr de la date de diagnostic. C'est bizarre,

on a l'impression que c'est quelqu'un qui avait été déjà diagnostiqué il y a deux, trois ans. »
À ce moment-là, on va renvoyer l'enquêtrice ou une autre enquêtrice sur une autre source d'information possible.

Pour mener ces enquêtes auprès des différentes organisations, les enquêtrices doivent négocier l'autorisation d'accéder aux éléments d'informations. Cette négociation auprès des agents fait intervenir tout un réseau d'éléments concernant l'accès à ces données de santé ; ces derniers pouvant alors entrer en opposition. Ainsi les initiatives telles que le Health Data Hub et les discours du « big data » ont alimenté une représentation des données comme *nouvel or noir* dont l'accès devrait être rémunéré ; cette acception peut entrer en contradiction avec une pratique antérieure d'échanges gratuits de données entre organismes publics, notamment pour des objectifs de recherche en santé publique.

Le travail d'enquêtes est identifié par les professionnels comme le plus consommateur en temps au sein des Registres. C'est précisément ce travail méticuleux qui participe à ce qui est identifié comme la qualité de l'ensemble des données et qui fonde la légitimité revendiquée en lien avec les missions premières des Registres. Ainsi, s'assurer de l'identité des personnes revient à démêler les éventuels problèmes d'homonymie entre des individus différents, mais aussi à tracer des individus dont la désignation peut changer : utilisation ou non d'un nom d'usage, prénom d'usage différent que celui mentionné en premier sur la pièce d'identité. Il s'agit aussi de vérifier des informations qui peuvent contenir des erreurs de transcription. Il peut également survenir des disjonctions entre la population étudiée et son territoire :

Resp. 1 : Par exemple, ce qui se passe. Il y a pas mal de gens, ils ont une chimiothérapie, ils ont une cure puis ils vont en vacances, ils font la cure ailleurs. Il y a pas mal de cas comme ça dans les deux sens, nous on le voit. Une fois, on a fait une étude sur les femmes qui avaient un cancer du sein, on repérait toutes les femmes, puis on avait des listes de chimiothérapies et en fait, on s'est rendu compte qu'il y avait des femmes qui habitaient très loin. Et quand on a regardé le dossier, elles étaient juste en vacances et comme elles avaient une cure de chimio à ce moment-là, elles l'ont faite à [Métropole du sud-ouest]. Alors inversement, on peut avoir des gens qui, pendant leurs vacances, la font ailleurs et là, ils vont nous échapper.

Comme le montre le verbatim suivant, établir le stade d'avancement du cancer demande plusieurs aller-retours entre l'encodage et l'enquête de terrain pour assurer l'instauration des données au fil des enregistrements produits par les encodeuses sur la base des preuves collectées par les enquêtrices.

Chercheur-e : Il y a un minimum requis.

Resp. 1 : Oui, c'est ça, on va le trouver. Après, pour l'échantillon stade, là où on insiste... parce que le problème des stades, de savoir quel est le stade, ça dépend de l'intensité avec laquelle vous avez cherché l'information. C'est-à-dire que quelqu'un qui a juste une radio, on ne risque pas de lui avoir... enfin, très peu de chance de lui avoir vu des métastases. Donc il y a un niveau de preuve, il y a ce qu'on a dit. Si on a dit « pas de métastase » sauf que si le niveau de preuve, c'est un ou zéro, ça veut dire qu'il peut très bien aussi

être métastatique. Parce qu'on n'a pas beaucoup cherché. Donc on est obligés d'avoir ces deux informations en même temps. C'est-à-dire que si on fournit le stade, on est obligés de fournir la fiabilité du stade. Et puis si on fournit le stade, on essaie de fournir aussi des informations sur le traitement parce que tant qu'à faire, ça ne coûte pas beaucoup plus cher d'avoir l'info sur... parce qu'elle est un peu au même endroit donc on prend aussi les informations sur le traitement.

Ainsi, les données « radio des métastases » et « traitement » sont en quelque sorte désencastrées de leur usage ancien dans le cadre du dossier patient, pour venir conforter l'identification des « cas » ainsi étayés par un faisceau d'indices.

Transformer en données

Les différents éléments recueillis par les enquêtrices sont encodés dans des bases de données selon un ensemble de préconisations et de standards émis à différents niveaux, tant nationaux qu'internationaux. Ces préconisations sont véhiculées à travers différents guides et notices édités par le Registre en question, le réseau national FRANCIM ou encore le CIRC (Centre International de Recherche sur le Cancer, agence de l'OMS spécialisée pour le cancer) ; ces différents guides et notices sont supposés s'articuler et être compatibles entre eux, selon un principe de granularité. En effet, les Registres et/ou le réseau national peuvent découper les classifications proposées par le niveau « supérieur » afin d'affiner les catégories, en particulier selon leurs axes de recherche. Les données ainsi instaurées sont susceptibles d'être recomposées, les faisceaux d'indices remis sur le chantier. Par ailleurs, des discontinuités sont toujours susceptibles d'apparaître. Le travail des données consiste de façon continue à construire de la continuité tout en l'ajustant à ces reconfigurations.

Resp. 1 : Un cancer, ça dure longtemps et ça change, ça se déplace un peu et donc à partir de quel moment on considère que c'est plus la même chose et qu'on doit réenregistrer ? Et comment on fait pour tenir compte d'une évolution ? Comment on gère l'évolution ? Ça, ce n'est pas résolu, encore. [...] Donc on essaye d'avoir des règles minimales et alors, là, nous on est passé à un peu plus complet que l'obligatoire. C'est-à-dire que nous on prend, effectivement, s'il y a deux seins, droite et gauche, on prend les deux. Et on ne prend un in situ que s'il n'y a pas d'invasif sur le même côté. Et ça, c'est à partir des cas de... en plus, il y a des débuts de règle d'enregistrement... enfin, si à partir d'un certain moment, on le fait, on ne peut pas le faire avant. Parce que si tous les gens se mettent à le faire pour des cas antérieurs, on sait plus à partir de quel moment on va considérer qu'on le fait... il faut qu'on ait des dates précises et en général, les dates ne portent pas sur la période où on travaille, mais sur la date d'incidence du cancer, ce qui est logique.

Transformer en données après avoir mené l'enquête, identifier des traces, instaurer les données s'avèrent des activités complexes et interdépendantes. Compte tenu de la complexité de ce travail, le choix a été fait dans l'organisation du sud-ouest d'une polyvalence, les salariés assurant en alternance les activités d'enquête et de codage pour constituer une connaissance partagée des difficultés liées à la codification et à

l'enquête. À l'inverse le Registre de l'ouest de la France a opté pour une spécialisation des tâches avec des postes de travail dédiés aux codages, même si les enquêtrices peuvent parfois aussi être chargées de coder. Le Registre de la façade méditerranéenne se situe à mi-chemin entre les deux organisations, puisque le codage est effectué à la fois par une secrétaire médicale et par les enquêtrices. Ces configurations organisationnelles traduisent différents compromis entre la spécialisation requise par activité, et la compréhension d'ensemble de la fabrique requise pour en assurer la consistance.

Assembler, désassembler, réassembler les données

Ces données sont par la suite recomposées, réagencées pour différentes finalités : calculs d'incidences, de mortalité, projets de recherche (au sein du Registre, en collaboration avec et/ou en réponse à des projets portés par d'autres organismes). Différents assemblages sont constitués au sein de chaque organisation-Registre pour faire connaître les résultats de son territoire : ensembles de données configurées par types de cancers, ou par groupes de patients atteints (en fonction des sexes, classes d'âge, comorbidités, etc.) (Fig. 3). Des assemblages et réassemblages sont mis en œuvre également pour l'envoi de sous-ensembles de données pour contribuer à d'autres bases de données (la base nationale FRANCIM, la base européenne des cancers, celle de l'OMS, etc.).

Cancer du poumon

Rapports standardisés de mortalité 2007 - 2014

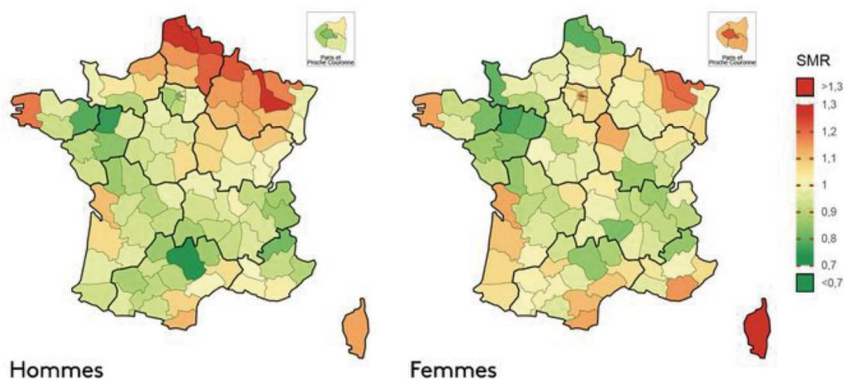


Figure 3 : La carte de la surmortalité des cancers en France entre 2007 et 2016 (rapports standardisés, source : Santé publique France).

Les organisations en charge des Registres sont chargées d'envoyer chaque année une extraction de leur base de données à la base nationale FRANCIM ; ces sous-ensembles de données comprennent l'ensemble des informations minimales que les Registres généraux doivent recueillir, à savoir la localisation de la tumeur, certaines caractéristiques, la date de diagnostic ainsi que le fondement de ce diagnostic. Les Registres n'envoient ni les éléments d'identification des patients ni les « sources » qui ont permis aux enquêtrices de collecter ces données. Ces données requises pour l'instauration des données doivent être effacées pour constituer une extraction transmissible, sachant que le respect de règles communes sur les modalités d'instauration doit permettre l'addition des données (les populations étant disjointes). Les agents des Registres opèrent alors une *re-brutification* des données qui vise ainsi à transformer « en données génériques, débarrassées de scories issues des activités situées qui s'en nourrissaient » (Denis 2018). La notion de brutification proposée par Denis vise à rendre compte de l'activité déployée au sein d'organisations pour construire des données supposément mobilisables pour tout usage. Ce que nous observons dans le cadre des Registres, c'est que ce qui relèverait de cette brutification est chaque fois spécifique, elle est relative, singulière, au sens où les combinaisons de données varient en fonction des finalités, et que de nouveaux assemblages peuvent requérir de nouveaux modes de brutification. La brutification est objet de tensions, de débats, de tractations récurrentes ; elle s'inscrit dans un contexte de relations de pouvoir, qui reposent notamment sur la mise en (in)visibilité par certains acteurs d'autres acteurs et actants afin de maintenir une forme d'autorité sur des données, des assemblages ou leur fabrique. Une des responsables de registre nous explique, avec un peu d'ironie, le voyage des données d'une administration à l'autre, voyage qui fait l'objet de tractations et facturations, et qui est aussi une transformation requise pour assurer la transmutation de données potentiellement identifiables vers des données dont l'anonymat est certifié et qui, de ce fait, vont pouvoir être transmises à un organisme tiers.

Resp. 1 : En fait, c'est de l'argent qui va à deux organismes. À l'INSEE, donc derrière, c'est Bercy, le Ministère des Finances, qui n'est pas facile à faire bouger. Et il y a des gens qui prennent la moitié de la somme, c'est l'INSERM. Ils prennent la moitié de la somme simplement car ils servent de tiers de confiance. C'est-à-dire que quand on envoie nos données, nous... moi, j'envoie les données, je m'appelle registre des cancers X. Donc ça veut dire que toute la liste des gens que j'envoie, ils ont un cancer. Je peux pas envoyer ça à l'INSEE puisque l'INSEE aurait un diagnostic. Donc je l'envoie à l'INSERM, l'INSERM « anonymise », c'est-à-dire enlève... met un numéro à la place de « registre des cancers du Tarn », et transmet. L'INSEE fait tourner sa moulinette, redescend sur l'INSERM, l'INSERM prend le numéro, remet « registre des cancers X » et me le renvoie. 150 000 euros. C'est bien payé. Et 150 000 pour dix minutes de temps machine pour l'INSEE parce que c'est pas long.

La fabrique des données est ainsi, de façon étroitement articulée, faite d'activités, de relations inter-organisationnelles, et affaire de politique, et d'enjeux socio-économiques.

Nous observons ainsi que les données au sein des Registres sont recomposées et réagencées en fonction des relations inter-organisationnelles et selon différents rapports d'autorité. Loin de s'additionner ou d'être mutuellement exclusifs ces différents agencements sont autant de « modes d'existence » des données (Simondon 2005, Lambotte 2017) instaurés au travers de négociations et de compromis inter-organisationnels qui mettent en jeu de vastes réseaux d'actants aux liens plus ou moins étroits, toujours susceptibles de discontinuité, toujours en reconstruction. Nous mesurons ainsi la complexité de la fabrique, et l'importance des savoirs professionnels requis pour cette instauration des données, pour en établir la pertinence et la cohérence. Les données sont instaurées en rapport avec une finalité première, ici celle de la veille sanitaire, qui en organise le territoire et les conditions de félicité. À l'opposé du discours dominant sur les « big data », notre travail ethnographique démontre que ces données ne sont pas « tout terrain » au sens où elles pourraient être mobilisées pour n'importe quel assemblage ; en revanche, elles peuvent servir à faire preuve, attester d'une évolution, établir des causalités, au-delà et distinctement de corrélations plus ou moins hasardeuses (Savy *et al.* 2020).

Conclusion

Après Leroi-Gourhan (1964) et Simondon (2005), Latour interroge l'aveuglement des 'Modernes' qui les conduit à n'accorder aucune attention à « l'art et la manière de déployer la technique » (Latour 2010 : 19). Cet aveuglement repose sur une conception qui conduit à voir dans « l'acte technique un transport sans transformation quand tout indique le contraire » (*ibid.* : 29). Une telle conception met la focale sur la fonction que l'objet technique est censé remplir, sur l'utilité, l'efficacité, en postulant « dans toute technique un transport d'efficacité à travers un outil parfaitement maîtrisé », conçu par « un fabricant qui possède dans sa tête une forme préalable qu'il applique à une matière inerte et informe » (*ibid.* : 21). Est-ce là comme le suggère Latour une pensée « hors-sol », sans prise avec l'art et la manière de (re)déployer la technique ? Une vue qui délègue en l'ignorant la mise en œuvre de toutes les opérations requises pour faire tenir ensemble des objets techniques toujours prêts à lâcher à la moindre occasion ?

Ce déni de prise en compte des modes d'existence des données est manifeste s'agissant de projets tels que le Health Data Hub. Des fonctions toutes plus considérables les unes que les autres sont mises en avant :

Les premiers axes de la feuille de route stratégique 2020-2022 sont d'ores et déjà fixés : il s'agira de décloisonner et d'améliorer la qualité de notre patrimoine de données, de renforcer l'usage par une offre de service attractive, de positionner la France comme un leader dans l'usage des données de santé, et de garantir la participation de la société civile au projet et l'exercice de l'ensemble de leurs droits. (document Health Data Hub. Offre de services 2020).

La priorité donnée à ces fonctions et à ce « transport d'efficacité » est apparue tellement impérative qu'un hébergeur a été retenu, en l'occurrence Microsoft Azure, toutes affaires cessantes, dès décembre 2019.

La focale a ainsi été portée sur la fonction visée. Le moyen de sa mise en œuvre est essentiellement ramené à la mise en place d'une plateforme de stockage, puisqu'il suffirait de transporter et d'additionner des données déjà là et intrinsèquement compatibles.

Or ce que l'exploration de l'instauration des données et de leur assemblage a permis de mettre en évidence, c'est l'ampleur et la complication de ce processus, qui donne existence aux données. Cette existence est toujours fragile, toujours susceptible de faire défaut en raison de manques d'éléments, de discontinuités des normes, de divergences entre les composants. L'instauration des données suppose de constituer de la continuité entre des indices dispersés, des traces conçues pour d'autres usages, et ainsi d'identifier un « cas », et de le suivre dans le temps. Les données instaurées sont toujours attachées à un « ensemble de personnes, de pratiques, de technologies, d'institutions qui les produisent, les maintiennent et les utilisent » (Courmont 2019). Loin d'être une réutilisation de données existantes par ailleurs, les instaurations de données nécessitent de détacher les éléments mobilisés dans les établissements de soin, dans les caisses d'assurances maladie, ... des pratiques qui leur sont associés et de les réaligner, de les rattacher à d'autres pratiques, d'autres institutions, d'autres définitions. Ces instaurations mobilisent tout un ensemble de savoirs professionnels et de pratiques afin de pouvoir faire sens de ces données. Les discours des *Big Data* et de l'open data tendent à invisibiliser tout ce travail, alors qu'il n'existe pas « de propriétés permanentes qui s'appliquent à toutes les données » (*ibid.*). L'instauration des données et leur assemblage mobilisent tout un réseau d'organisations, d'agents, d'objets et de systèmes techniques, réseau qui est toujours en reconstruction, soumis à des tensions, des rapports de force, des questions d'autorité et de légalité.

Bibliographie

- COULON Alain, 2014, *L'ethnométhodologie*. Paris, Presses Universitaires de France, « Que sais-je ? ». ISBN : 9782130634935.
- COURMONT, Antoine, 2019, « Ce que l'open data fait à l'administration municipale. La fabrique de la politique métropolitaine de la donnée », *Réseaux*, 218 : 6. 77-103.
- DENIS Jérôme, 2018, *Le travail invisible des données. Éléments pour une sociologie des infrastructures scripturales*, Paris, Presses des Mines.
- DENIS Jérôme, GOËTA Samuel, 2017, « Rawification and the careful generation of open government data », *Social Studies of Science*, 47: 5. 604-629.
- DESROSIÈRES Alain, 2008, *Gouverner par les nombres. L'argument statistique II*, Paris, Presses des Mines.
- GROSJEAN Sylvie, 2013, « Une approche microethnographique et multi-située en organisation », *Revue internationale de psychosociologie et de gestion des comportements organisationnels*, (Supplément), 155-177.
- LAMBOTTE, François, 2017. « Modes d'existence de la mémoire organisationnelle : ce que le numérique nous donne à voir et à explorer ». Les Décades de Cerisy : quelles communications, quelles organisations à l'ère du numérique ?
- LATOUR Bruno, 2007, *Changer de société, refaire de la sociologie*. Paris, La Découverte.
- LATOUR Bruno, 2010, « Prendre le pli des techniques », *Réseaux*, 163 : 5. 11-31.
- LATOUR Bruno, 2014, « Reflections on Etienne Souriau's Les différents modes d'existence », *The speculative turn; Continental Materialism and Realism*. 304-333.
- LEHMANS Anne, 2017, « Données ouvertes et redéfinition de la culture de l'information dans les organisations », *Communication et organisation*, 51, 15-26.
- LEROI-GOURHAN André, 1964, *Le geste et la parole. Tome 1 : Technique et Langage*. Paris, Sciences d'Aujourd'hui, Albin Michel.
- MARTIN-SCHOLZ Anja, 2019, « L'inter-organisationnel comme ressource entre ambiguïté et autorité », Colloque international Org&Co. 28-29 mars 2019, Bordeaux.
- MAYÈRE Anne, 2018, « L'utopie de l'organisation numérique », dans L. Bonneville, S. Grosjean et A. Mayère, *Les utopies organisationnelles*, Éditions ISTE, 41-66.
- MONDOUX André, MENARD Marc, 2018, « Big data, circuits marchands et accélération sociale », dans A. Mondoux et M. Ménard (dir.), *Big Data et société*, Presses de l'Université du Québec, 63-83.
- NICOLINI Davide, 2009, « Zooming In and Out: Studying Practices by Switching Theoretical Lenses and Trailing Connections », *Organization Studies*, 30 : 12. 1391-1418.
- PUIG DE LA BELLACASA Marie, 2011, « Matters of care in technoscience: Assembling neglected things », *Social Studies of Science*, 41(1), 85-106.

- SAVY Nicolas, MAYÈRE Anne, MARTIN-SCHOLZ Anja et LAMBOTTE François, 2020, La fabrique des données à l'épreuve des programmes de big data, *Actualité et dossier en santé publique*, 112, 27-29.
- SIMONDON Gilbert, 2005, *L'invention dans les techniques*. Cours et conférences, éditions du Seuil.
- SUCHMAN Lucy, 1993, « Do categories have politics? The language/action perspective reconsidered », *Proceedings of the Third European Conference on Computer-Supported Cooperative Work 13-17 September 1993, Milan, Italy ECSCW'93*.1-14.
- WATHELET Emmanuel, 2015, *Autorité et auteurités émergentes dans un environnement organisationnel non-hiérarchisé en ligne : le cas de la construction de règles sur Wikipédia*, thèse, Université catholique de Louvain.

Biographies des auteurs

Anja Martin-Scholz, Maître de conférences, UMR5044 CERTOP, Université Paul Sabatier Toulouse III. Ses recherches, inscrites dans le champ de la communication organisationnelle, interrogent le travail des données en ce qu'il prend forme dans et entre des organisations, dans un contexte où les collaborations inter-organisationnelles se multiplient, en explorant les spécificités de ces questions dans le champ de la santé et des organisations de santé.

Anne Mayère, Professeure émérite, UMR5044 CERTOP, Université Paul Sabatier Toulouse III, Chercheure en sciences de l'information et de la communication, ses recherches s'inscrivent dans le champ de la communication organisationnelle, sur le terrain des organisations de santé, et des transformations des relations soignants - patients.

François Lambotte, Professeur, LASCO, Université Catholique de Louvain. Il a créé le Social Média Lab (SML) de l'UCLouvain (www.socialmedialab.be) dont les recherches questionnent l'utilisation des médias sociaux en contexte professionnel. Ses recherches s'inscrivent dans le domaine de la communication organisationnelle, plus particulièrement sur la thématique de l'écriture collective du changement organisationnel, et sur la fabrique des métriques des médias sociaux (*social media analytics*).