

Instauration et assemblage des données en santé : le cas des Registres des cancers

Health Data Elaboration and Assembling: the Case of the Cancer Registries

Anja MARTIN-SCHOLZ, Jeune docteure, UMR5044 CERTOP, Université Paul Sabatier

Toulouse III, anja.martinscholz@gmail.com

Anne MAYÈRE, Professeure, UMR5044 CERTOP, Université Paul Sabatier Toulouse III,

anne.mayere@iut-tlse3.fr

François LAMBOTTE, Professeur, LASCO, Université Catholique de Louvain,

francois.lambotte@uclouvain.be

Mots-clés : santé ; données ; bases de données ; autorité ; instauration ; assemblage

Keywords: health; data; database; authority; elaboration; assembly

Résumés : (500 signes)

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.

Abstracts: (500 signs)

The rise of discourses promoting “Big Data” suggests that gathering and combining a sufficient amount of “data” could provide a comprehensive view of phenomena. Building on a current research in the field of organizations in charge of producing Cancer Registries, we propose to specify the process of data *instauration* in conjunction with their assemblies. Thus will be characterized what underlies the *modes of existence* of data and their heterogeneous and unstable assemblies. Going beyond the displayed purposes, the aim is to better understand what is at work in the data factory, through zooms in and out. This identifies issues in terms of knowledge but also in political, economic and social terms.

# Instauration et assemblage des données en santé : le cas des registres du cancer

Anja MARTIN-SCHOLZ, Anne MAYÈRE, François LAMBOTTE

Evaluer 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 contagion du Covid-19 est plus qu'emblématique en la matière. Le manque d'homogénéité des définitions, et/ou d'exhaustivité du comptage à partir de ces définitions, soulève le problème d'une impossible comparaison entre pays, quant aux chiffres de morbidité et de mortalité, et quant aux courbes qui suivent l'évolution de la maladie.

La création du Health Data Hub en 2019, 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](http://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.

*Avec l'allongement de la durée de vie et les progrès de la médecine, les contextes cliniques des patients se complexifient. Par exemple, les personnes âgées souffrent de plus en plus d'une pluralité de pathologies et sont amenées à être suivies par différents professionnels de santé et structures médicales. Le praticien doit alors arbitrer pour offrir le meilleur traitement tout en prenant garde à la qualité de vie du patient, notamment lorsque les traitements sont lourds et admettent d'importants effets secondaires. Analyser 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[...] A 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. L'enjeu est de taille pour la prise en charge du cancer, puisque cette pathologie est encore responsable en France de 30% des décès (plus de 160 000 par an) avec 400 000 nouveaux cas par an et un coût annuel de 15 milliards d'euros. (Health Data Hub, Mission de préfiguration, 12/10/2018, p.9-10 publié sur [solidarites-sante.gouv.fr](http://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.

Ce faisant, ils minimisent et invisibilisent le travail complexe des données requis tant pour leur construction que pour leur intelligibilité (Denis & Goëta, 2017).

*Aujourd'hui, l'élaboration de ces algorithmes requiert une annotation manuelle des signaux ou images par des spécialistes, mais on peut imaginer que le chaînage avec d'autres sources de données pourrait faciliter la qualification des données et dynamiser l'innovation en la matière. D'autres tâches peuvent être facilitées par des algorithmes ; plusieurs expériences visent à mobiliser les comptes-rendus médicaux pour automatiser le codage des actes et diagnostics à l'hôpital dans le cadre de leur remboursement par l'Assurance Maladie... (Health Data Hub, Mission de préfiguration p.10)*

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 acceptions 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 communication s'appuie sur une recherche en cours sur le terrain d'organisations en charge de la production des Registres des cancers de Loire-Atlantique/Vendée d'une part et du département du Tarn d'autre part. Les Registres des Cancers sont des organisations, le plus souvent sous statut d'association à but non lucratif, 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). 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 soin, les campagnes de prévention ou encore l'influence des inégalités sociales sur l'incidence et/ou la survie. Les Registres sont ainsi évalués sur la qualité des données, mais aussi et surtout sur les publications effectuées en leur sein.

Les deux organisations enquêtées sont des Registres généraux ayant une portée départementale, c'est-à-dire qu'ils enregistrent tous les cas de cancer d'individus vivant dans le département en question, quel que soit l'âge ou l'organe impacté et visent à recueillir un minimum d'information selon des standards internationaux.

Les Registres généraux peuvent aussi développer des programmes et/ou des axes de recherche particuliers pour lesquels ils vont enregistrer des informations complémentaires.

Une des principales préoccupations des chercheurs de ces Registres est de s'assurer de l'exhaustivité des cas enregistrés pour une zone donnée, en l'occurrence à l'échelle d'un département. En effet, pour pouvoir faire des comparaisons d'incidence, savoir si le nombre de cas de cancers augmente ou diminue d'une année sur l'autre, les chercheurs contrôlent qu'ils mesurent bien les mêmes choses. Pour ce faire, les définitions sur lesquelles sont fondées l'enregistrement des cas se basent sur des critères partagés internationalement et nationalement. Ainsi, la définition de ce qu'est un cancer est fondée sur un ensemble de critères définis au sein du Centre International de Recherche sur le Cancer (CIRC) de l'OMS – d'une cellule avec un fonctionnement « anormal » à la tumeur maligne, il existe tout un continuum. La définition de la zone où sont enregistrés les cancers est également très importante, pour s'assurer qu'elle reste constante dans le temps et pour établir la population concernée de façon maîtrisée et homogène. La définition retenue est celle utilisée par l'Insee pour mesurer la population d'une zone géographique donnée, à savoir toutes les personnes vivant plus de six mois dans le département en question.

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 sont bien des cancers selon la définition de l'OMS. Pour ce faire, les agents des Registres multiplient les sources de notifications de cas de cancer et les croisent. En effet, aucune source n'est exhaustive. Ainsi, tous les cancers ne font pas obligatoirement l'objet d'une biopsie, ni ne font forcément l'objet d'une hospitalisation.

L'investigation de terrain a pris la forme d'entretiens semi-directifs mené auprès de médecins épidémiologistes de ces Registres (4 entretiens de 5h), ainsi qu'auprès de chercheurs épidémiologistes spécialisés dans l'utilisation de ces données (3 entretiens de 5h10'). Par ailleurs, un ensemble de documents ont été rassemblés, tels que le manuel servant en appui au travail de codification, ou la fiche de collecte d'informations.

## **Une fabrique faite d'instauration et d'assemblage**

Au travers de la présentation des Registres nous pouvons déjà percevoir que l'identification des cas de cancers est beaucoup plus complexe que ce que le terme « recueil » laisse supposer. Cette première description fait poindre tout un ensemble d'actants et de formes organisationnelles mobilisés au travers de cette fabrique de données. Pour étudier plus en avant cette dernière nous proposons de faire un double mouvement de « zoom avant/zoom arrière » (Nicolini, 2009 ; Grosjean, 2013). En effet, les activités situées au sein des Registres sont connectées à un ensemble d'éléments situés dans un « ailleurs » (Latour, 2007). Il s'agit à la fois de prêter suffisamment 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). Nous proposons par la suite de nous intéresser aux différentes « opérations » (Denis & Goëta 2017) qui organisent le travail des Registres de cancer en faisant à la fois ce mouvement de « zoom avant » de description des pratiques et de « zoom arrière » en montrant comment ces activités sont connectées à d'autres lieux et mobilisent des ressources organisationnelles. A l'instar de Denis et Goëta qui, 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 » (2017 p.4), nous proposons d'analyser l'agencement de ce réseau d'actants pluriels au rythme des opérations d'instauration et d'assemblage, dont nous allons explorer les dimensions et les articulations.

### **L'instauration des données**

Nous mobilisons la notion d'*instauration*, en nous fondant sur la proposition de Denis (2018) et de Lambotte (2017), qui s'appuient eux-mêmes sur Latour (2010 ; 2014). Latour retravaille la notion d'instauration à partir de Souriau en montrant comme 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., p18). Latour se démarque de la conception constructiviste qui veut que « les faits soient construits », dans notre cas, les données ne seraient que des « construits ». Le savant ne construit pas les faits à partir de rien, ni les enquêtrices s’agissant des données, pas plus à l’inverse qu’ils ne les « collectent ». 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., p 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 le seul résultat. Par ailleurs, la focale n’est pas sur les acteurs, mais sur les actions (Lambotte, 2017) et sur les relations, sur lesquelles il s’agit « d’enquêter de façon sérieuse » (Latour, 2010, p17).

Les données ne sont ni *déjà-là*, ni construites comme ex-nihilo, mais *instaurées*. Denis a déjà montré dans ses recherches sur les programmes d’open data que l’idée que les données attendent 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 données, il faut donc que les agents des Registres les *instaurent*, 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 rassemblant des cas toujours singuliers en autant de catégories établies par des définitions internationales.

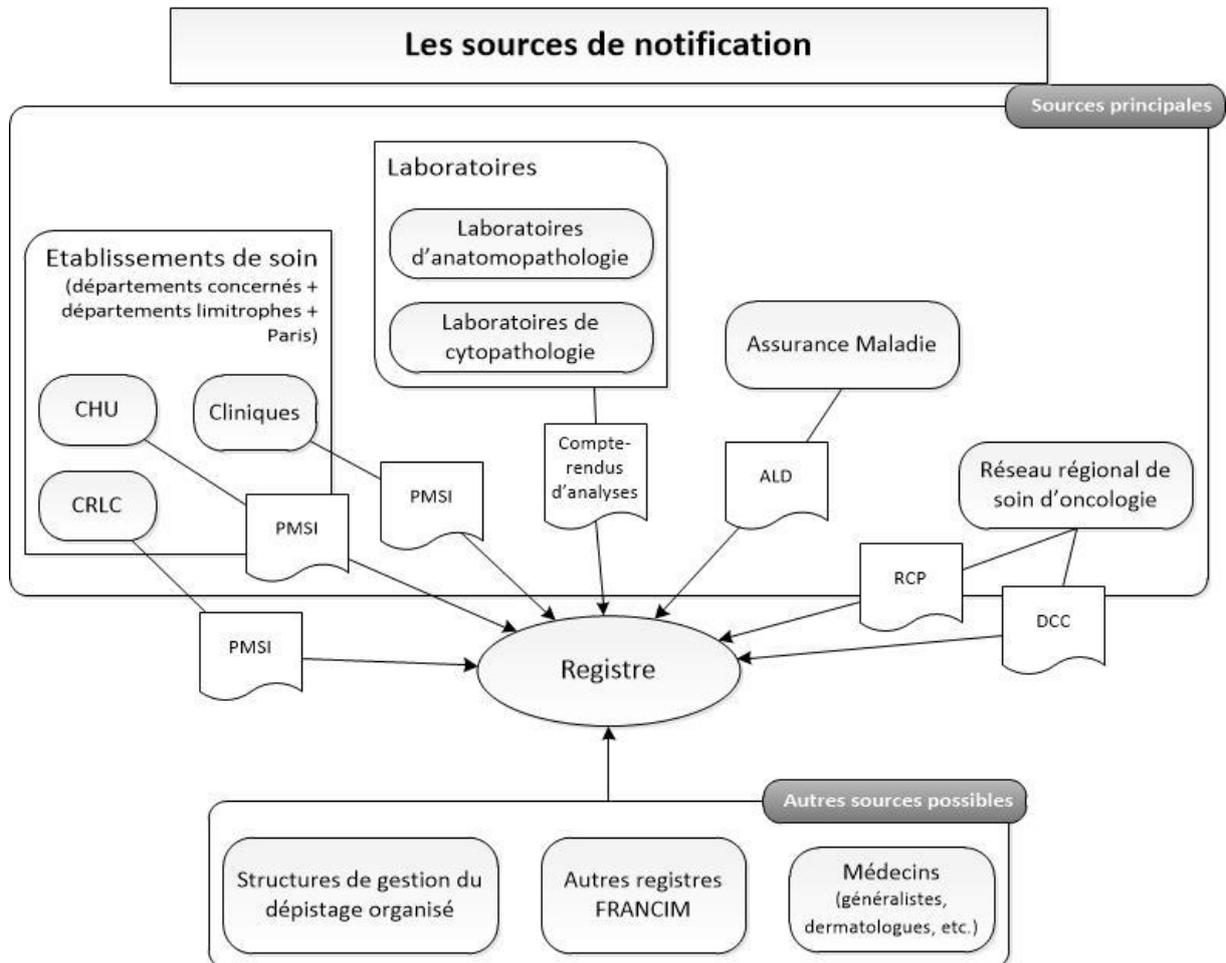
Il peut y avoir différents registres d’instauration entre les données considérées par les enquêtrices comme étant « de fonctionnement », celles qui permettent des traitements, celles qui se prêtent à des agrégations. Cette instauration des données suppose d’identifier des traces, de mener l’enquête, de transformer en données par la codification. Elle est de façon étroitement liée affaire d’assemblage. 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.

#### Identifier des traces

Les organisations dénommées Registres ont pour rôle d’identifier tous les nouveaux « cas » de cancers sur le territoire couvert, à savoir ici les départements de Loire-Atlantique et de Vendée pour le Registre de Loire-Atlantique/Vendée et celui du Tarn pour le Registre du

Tarn. Afin d'enregistrer les cas de cancers, les enquêtrices vont solliciter différents organismes, des « sources de notification », afin de recueillir auprès d'eux différents documents qui leur permettront d'identifier des indices de cas « potentiels » de cancers. Des « sources » sont identifiées, documentées, des liens sont établis avec des interlocuteurs dans les établissements de santé sur le territoire pour obtenir à échéance régulière (au moins annuellement) des enregistrements sur les hospitalisations. Des formats de fichiers et d'échange sont expérimentés, routinisés, reconstitués au gré des transformations.

Comme le figure le schéma suivant, l'élaboration d'un registre repose sur de multiples sources de notification (PMSI, Comptes-rendus d'analyses, ALD, RCP, DCC) qui, elles-mêmes, dépendent d'une multitude d'organisations (hôpitaux, laboratoires, Assurance maladie, etc.).



Loin d'être aussi facile que ce que les termes « recueil » ou « notification » le laisseraient à penser, l'accès à ces enregistrements, à des traces, se fait par des médiations plurielles, des mises en relation, 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 mises en relation locales et des autorisations de mises en lien requérant l'intervention d'instances nationales voire internationales. Ainsi, sachant que les documents en question relèvent de la définition des « données de santé », leur accès fait à la fois l'objet de négociations entre les enquêtrices des Registres et les personnes en charge dans diverses organisations de la tenue et de la gestion d'enregistrements, mais aussi de discussions à un niveau plus global, afin d'obtenir les autorisations requises.

MEL : *On rentre pas dans les cases de recherche-j'sais-pas-quoi.*

L'accès aux traces se double de la mémorisation des cheminements par lesquels elles ont été constituées de façon à attester de leur fiabilité, et à pouvoir si nécessaire les emprunter à nouveau pour compléter ces traces. Il s'agit ainsi d'identifier ce qui dans les termes du métier d'enquêtrices sont appelés des « données de fonctionnement » pour pouvoir mener l'enquête sur les cas et leur évolution, à savoir identifier les acteurs (médecins, secrétariats médicaux et/ou centres de soin) susceptibles de repérer et de donner des nouvelles des patient-e-s concernées. Ces « données de fonctionnement » peuvent aussi permettre de mener des études ultérieures si besoin. Selon les spécialisations de recherche de l'organisation ou les programmes de recherche, les caractérisations visées sur certains sous-ensembles de patients peuvent être étendues ou les règles de codification peuvent évoluer, ce qui peut nécessiter de revenir sur les traces pour les compléter.

### Mener l'enquête

Il n'existe donc pas de liste de données pré-identifiées qui attendent d'être mobilisées au sein des organismes « sources de notification ». Au contraire, les données doivent être *instaurées* en tant que données pour les Registres de cancer. En effet, avant d'être des données pour la veille sanitaire sur le cancer et/ou la recherche épidémiologique sur cet ensemble de maladies, les éléments, les 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. Leur identification en tant que traces mobilisables pour instaurer des cas recensés nécessite tout un travail de négociation entre acteurs locaux et globaux. En cela, à la différence de Denis qui investit l'instauration des données issues d'une source unique, la fabrique du registre repose sur une multitude de sources permettant de transformer des indices dispersés en données construisant un « Registre des cancers ».

Pour étalonner ce qui va contribuer à les instaurer en données du Registre, les enquêtrices se munissent de deux définitions : quelle est la population à laquelle elles s'intéressent et quels sont les critères définissant un « nouveau cas de cancer ». Nous l'avons déjà vu, la définition de la population s'appuie sur une définition de l'Insee, tandis que celle des nouveaux cancers sur des critères édictées par l'OMS, soit deux autorités et des sous-ensembles de normes et procédures associés.

Une fois identifié-e-s des patient-e-s, 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 récives 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 étudier les cancers au sein des Registres. En effet, savoir quels soins ont été nécessaires, les proportions de récives pour chacun des cancers ou la capacité de tel ou tel cancer à faire des métastases dans tel ou tel organe sont des questions légitimes de recherche. Néanmoins les Registres s'inscrivant dans une optique de veille sanitaire, il s'agit pour leurs chercheurs d'identifier en priorité les facteurs de risques et les variations « anormales » du nombre de malades, en estimant que les évolutions d'un cancer donné ont la même cause, s'appuient sur les mêmes facteurs que le cancer initial.

Le travail des enquêtrices consiste ainsi en un long et méticuleux travail consistant à créer une nouvelle fiche, vérifier les informations déjà contenues dans une fiche et de compléter cette dernière. Ainsi, elles détectent et rassemblent 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 et les cliniques.

Ce travail d'enquête repose sur un chaînage précis d'éléments initialement construits pour d'autres objets en autant d'indices susceptibles d'instaurer la validité et la caractérisation de chaque cas, et de son classement adéquat.

MET : *Un cas standard, ça [ne] prend pas énormément de temps. Le problème, ce sont les gens qui sont arrivés à un endroit, ça pas marché, ça, ça prend énormément de temps.*

**ITW : *Et ça se passe comment pour ces cas-là ?***

MET : *Alors nous, ça part et ça revient. Quand les enquêtrices ramènent les cas, elles les donnent à la personne qui code. Ça peut être l'une ou l'autre, vu qu'on croise les choses. 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. » A ce moment-là, on va renvoyer l'enquêtrice ou une autre enquêtrice sur une autre source d'information possible. En plus, nous, on essaie d'optimiser donc on n'enquête pas partout systématiquement. On va sur la source principale d'abord et si là, il n'y a pas assez, on va relancer les autres sources. On serait une énorme structure, ce serait compliqué, mais là, les enquêtrices se parlent entre elles donc s'il y en a une qui revient, elle sait aussi qu'il n'y a pas assez, elle s'en rend compte. Donc elle dit « non, celle-là, tu vois, j'ai trouvé que ça donc il faut demander à machin d'aller dans tel hôpital parce que je n'ai pas assez ».*

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 la mise en avant des dites « 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 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 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.

*MET : 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 Toulouse. Alors inversement, on peut avoir des gens qui, pendant leurs vacances, la font ailleurs et là, ils vont nous échapper. Sauf effectivement si on avait une base nationale, ce qui n'est pas le cas encore pour l'instant, puisque nous on n'a que la base des PMSI qu'on a demandés. Donc je peux savoir éventuellement que quelqu'un qui devait avoir une chimiothérapie à Toulouse l'a eue à Auch parce que j'ai reçu... enfin, si j'ai reçu le dossier de là-bas. Mais s'il l'a eue à Saint-Dié ou dans le nord de la France, je ne vais pas le savoir pour l'instant. Alors c'est sûr que si j'avais tous les remboursements, ce serait mieux. Tous les remboursements et séjours hospitaliers.*

De même, s'assurer qu'il s'agit bien d'un « nouveau cancer » suppose de décortiquer l'ensemble des éléments à disposition pour vérifier qu'il s'agit bien d'un cas de cancer au sens de l'OMS et que le signalement identifié ne tient pas à l'évolution d'un cancer préexistant. Ces aiguillages dans la codification peuvent changer en fonction des types de cancers et dans le temps, certaines avancées de recherche ou les partages d'expérience entre professionnels des registres pouvant amener à distinguer des sous-catégories ou introduire de nouvelles cartographies des cancers (ainsi quand ce qui était identifié comme un type de cancer est requalifié en récurrence d'un autre « cancer primitif »).

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

**ITW : Il y a un minimum requis.**

*MET : 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 brutifiées, désencastrées de leur usage ancien dans le cadre du dossier patient, pour créer de l'intelligibilité dans le contexte d'usage du registre, et 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 ; 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, et le travail des données consiste de façon continue à

*MET : Il y avait beaucoup de problèmes de mémoire. Mais c'est un gros problème, de savoir... 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. Parce que vous pouvez avoir des gens, ils vont renseigner tous les jours en disant « aujourd'hui, ça n'a pas bougé, aujourd'hui, il s'est passé ça », des gens qui vont encrasser le système et c'est très difficile d'avoir des règles. 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. Après, il y a les cas particuliers... avec des exemples. En fait, c'est fait pour aider les gens qui ne connaissent pas trop non plus. En même temps, il faut quand même connaître un petit peu. Avoir un minimum de culture sur le sujet.*

construire de la continuité tout en l'ajustant à ces reconfigurations.

Compte tenu de la complexité de ce travail, le choix a été fait dans l'organisation tarnaise d'une polyvalence, les salarié·e-s 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 Loire-Atlantique 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.

## **Assembler, désassembler, réassembler les données**

Ces données sont par la suite recomposées, réagencées pour différents usages : 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 en interne pour différents usages : ensembles de données configurées par types de cancers, ou par groupes de patients atteints (en fonction des sexes, classes d'âge, comorbidités, etc.). Ces assemblages et réassemblages sont mis en œuvre également pour l'envoi de sous-ensembles de données visant à « nourrir » d'autres bases de données (la base nationale FRANCIM, la base européenne des cancers, celle de l'OMS...).

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

*MET : En fait, c'est de l'argent qui va à deux organismes. A 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 du Tarn. 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 du Tarn » et me le renvoie. 150000 euros. C'est bien payé. Et 150000 pour dix minutes de temps machine pour l'INSEE parce que c'est pas long.*

transmises à un organisme tiers.

Les Registres sont chargés 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 ; il s'agit là selon la dénomination indigène des « données de fonctionnement » requises pour l'instauration des données mais qui doivent être ensuite 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). Si la notion de brutification proposée par Denis s'inscrit dans un contexte visant à construire des données supposément mobilisables pour tout usage, dans le cadre des Registres, cette re-brutification est chaque fois spécifique, elle est relative, singulière, au sens où les

combinatoires de données varient en fonction des finalités, et que de nouveaux assemblages peuvent requérir de nouveaux modes de brutification. Cette brutification est objet de tensions, débats, tractations récurrentes ; elle s'inscrit dans un contexte de jeu de relations de pouvoir et de mise en (in)visibilité afin de maintenir une forme d'autorité sur certaines données.

En dehors de ce fonctionnement routinier, les Registres peuvent être aussi ponctuellement sollicités par des chercheurs, de ce dernier ou extérieurs, soit pour opérer des calculs particuliers sur des données constituées, soit pour accéder à un sous-ensemble de données anonymisées selon le format requis après avoir rempli une demande présentant le protocole de recherche selon les argumentaires requis.

Nous observons ainsi que les données aux seins des Registres se voient 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 1987, 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.

## **Conclusion**

Après Leroi-Gourhan et Simondon, 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, p. 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. p. 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. p21). Est-ce là comme le suggère Latour une pensée « hors-sol », sans prise avec l'art et la manière de 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 de l'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. Il s'agit également et conjointement de constituer et maintenir des assemblages de données hétérogènes et toujours susceptibles de diverger. Par l'assemblage, le désassemblage, le réassemblage des données, se construisent et se reconstruisent des modes d'intelligibilité. 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.

Si le grand projet du Health Data Hub, à peine lancé, ressemble déjà à ces grandes arches numériques sans lendemain, cela semble être en bonne part lié à cette ignorance radicale, « au carré », des modes d'existence tant des données que de leur assemblage.

## **Bibliographie**

Denis, J. (2018). *Le travail invisible des données. Éléments pour une sociologie des infrastructures scripturales*. Paris, Presses des Mines

Denis, J. & Goëta, S. (2017). Rawification and the careful generation of open government data. *Social Studies of Science*, n°47(5), p. 604–629.

- Desrosières, A. (2008). *Gouverner par les nombres. L'argument statistique II*. Paris, Presses des Mines.
- Grosjean, S. (2013). Une approche microethnographique et multi-située en organisation. *Revue internationale de psychosociologie et de gestion des comportements organisationnels*, (Supplément), p. 155-177.
- Lambotte, F. (2017, July). Making Social Media Activity Analytics Intelligible for Oneself and for Others: A “Boundary Object” Approach to Dashboard Design. *International Conference on Human Interface and the Management of Information*. Springer, Cham. p. 112-123.
- Latour, B. (2007). *Changer de société, refaire de la sociologie*. Paris, La Découverte.
- Latour, B. (2010). Prendre le pli des techniques. *Réseaux*, vol. 163(5), p. 11-31.
- Latour, B. (2014). Reflections on Etienne Souriau's Les différents modes d'existence. *The speculative turn; Continental Materialism and Realism*. p. 304–333.
- Martin-Scholz, A. (2019) *L'inter-organisationnel comme ressource entre ambiguïté et autorité*. Colloque international Org&Co. 28-29 mars 2019. Bordeaux.
- Mayère A., 2018, « L'utopie de l'organisation numérique », dans L. Bonneville, S. Grosjean, A. Mayère, Les utopies organisationnelles, Editions ISTE, p.41-66.
- Nicolini, D. (2009). Zooming In and Out: Studying Practices by Switching Theoretical Lenses and Trailing Connections. *Organization Studies*, n°30(12), p. 1391–1418.
- Suchman, L. (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*. p.1-14
- Wathelet, E. (2015). *Autorité et auteurités émergentes dans un environnement organisationnel non-hiérarchisé en ligne: le cas de la construction de règles sur Wikipédia*. Thèse de doctorat.