
L'artisanat des données à l'ère de la *datafication*

François Lambotte et Kevin Carillon



Édition électronique

URL : <https://journals.openedition.org/communication/22019>

DOI : 10.4000/15boc

ISSN : 1920-7344

Éditeur

Université Laval

Ce document vous est fourni par Université catholique de Louvain



Référence électronique

François Lambotte et Kevin Carillon, « L'artisanat des données à l'ère de la *datafication* », *Communication* [En ligne], Vol. 42/2 | 2025, mis en ligne le 11 décembre 2025, consulté le 12 décembre 2025. URL : <http://journals.openedition.org/communication/22019> ; DOI : <https://doi.org/10.4000/15boc>

Ce document a été généré automatiquement le 11 décembre 2025.



Le texte seul est utilisable sous licence CC BY-NC-ND 4.0. Les autres éléments (illustrations, fichiers annexes importés) sont susceptibles d'être soumis à des autorisations d'usage spécifiques.

L'artisanat des données à l'ère de la datafication

François Lambotte et Kevin Carillon

- 1 Les discours et les imaginaires associés à la *datafication* de la société, c'est-à-dire la propension à considérer chaque aspect du réel comme une source potentielle de données et, partant, de valeur économique (Mayer-Schönberger et Cukier, 2013), mettent en permanence l'accent sur la dimension quantitative des données. Que l'on parle des *big data* ou des immenses masses de données ayant servi à l'entraînement des IA génératives, tout semble concourir à l'idée que nos régimes contemporains de production et d'exploitation de données se suffisent d'une accumulation toujours plus étendue et intensive pour exister. Pour autant, les spécialistes de la donnée, comme la littérature scientifique, ont depuis longtemps battu en brèche cette idée en mettant en avant la qualité des données comme un enjeu crucial (Cai et Zhu, 2015 ; Dimas, Goldkind et Konrad, 2023). Dans ce contexte, une donnée pertinente est donc une donnée sélectionnée, altérée ou façonnée. Dès lors, loin de l'idée véhiculée par l'appellation « donnée brute » (celle d'une donnée immanente, disponible et dont il suffirait de se saisir), les données qui soutiennent ces régimes apparaissent comme des construits (Gitelman, 2013 ; Barrowman, 2018).
- 2 Or, qui dit construit dit également travail de construction. Dans ce contexte, il est symptomatique de constater qu'une grande partie de ce travail est largement invisibilisée (Denis, 2018 ; Gray et Suri, 2019 ; Rothschild *et al.*, 2022). Invisibilisée, tout d'abord, par les longues chaînes de médiations sociotechniques qui composent les assemblages de données, dont les composants sont dispersés dans le temps et l'espace. Invisibilisée, ensuite, par la nature souvent répétitive, laborieuse et peu valorisée des tâches assurées par des travailleureuses humaines permettant la production des données. Invisibilisée, enfin, par la recherche elle-même, qui s'intéresse davantage au centre des régimes de production et d'exploitation (Muller *et al.*, 2019 ; Stelmaszak, 2022) qu'aux marges et aux périphéries au sein desquelles ces tâches ont lieu.
- 3 Cet article vise à mettre en lumière et à caractériser une partie de ce travail de construction en le saisissant à travers le concept d'artisanat des données (*data crafting*).

Relevant du champ plus large du travail des données, entendu ici comme l'ensemble des activités liées « à la création, la collecte, la gestion, la conservation, l'analyse, l'interprétation et la communication des données » (Bossen *et al.*, 2019, p. 466), l'artisanat des données permet de désigner un ensemble de pratiques qui échappent aux pures logiques d'exploitation ou de précarisation (Casilli, 2015 ; Irani, 2015 ; Gray et Suri, *op. cit.*) dans le travail des données. Ainsi met-il davantage l'accent sur les savoir-faire, le soin apporté aux données et les communautés de pratique qui façonnent les données dans des contextes situés. Autrement dit, là où la littérature a surtout décrit la face sombre du travail des données, le concept d'artisanat ouvre la possibilité de penser des pratiques qui sont certes laborieuses et souvent invisibilisées, mais qui mobilisent également des formes d'expertise, d'engagement collectif et de sens pour les travailleur·euse·s concerné·es.

- 4 Pour développer notre proposition, nous proposons, tout d'abord, de passer en revue la manière dont la littérature scientifique caractérise l'artisanat, puis de dresser un panorama des études sur le travail des données afin de mettre en évidence l'absence d'une perspective proprement artisanale dans ces travaux. En vue d'illustrer ce que nous entendons précisément par artisanat des données, nous mobiliserons ensuite deux terrains ethnographiques. Le premier est une étude menée dans des organismes responsables de la production de registres du cancer (RC) dans deux départements français, tandis que le second concerne la communauté Stack Overflow (SO) sur Stack Exchange, un forum en ligne fondé par des développeurs de logiciels, devenu une référence dans le domaine de la programmation. À travers ces données empiriques, nous montrerons en quoi le travail des acteur·rice·s de ces deux terrains s'apparente à de l'artisanat des données tout en en mettant en exergue ses caractéristiques. Enfin, nous analyserons la manière dont les artisan·es des données s'appuient, précisément, sur le caractère artisanal de leurs pratiques pour faire sens de la massification des données et des potentialités de l'IA dans leurs domaines respectifs.
- 5 Cet article contribue ainsi à la littérature sur le travail des données à l'ère de la *datafication*. D'abord, il propose un concept théorique original pour penser et analyser un ensemble de pratiques qui s'inscrivent dans certains contextes de production. En mettant en valeur le travail des artisan·es de la donnée, il invite également à une prise en compte accrue des tâches invisibles, et pourtant essentielles, qui participent au fonctionnement et à la persistance de nos régimes de production et d'exploitation des données. Enfin, en soulignant les compétences, les savoir-faire et les valeurs qui sous-tendent ces pratiques, l'artisanat des données met en évidence la dimension à la fois éthique et politique de certaines formes de travail des données, souvent occultée par les discours technocentrés sur la *datafication* et les progrès de l'IA.

L'artisanat au croisement du développement d'habiletés, du *care* et de la matière

- 6 Comment la littérature scientifique appréhende-t-elle l'artisanat ? Cette question constitue le point de départ de la présente section, qui explore les principales dimensions mises en évidence par la littérature pour caractériser ce phénomène. Les recherches sur l'artisanat s'articulent ainsi autour de trois dimensions complémentaires : (1) le développement d'habiletés au sein de communautés de pratique, (2) le soin apporté aux choses et (3) le rapport à la matière travaillée. En

croisant ces trois dimensions, la littérature offre une compréhension fine de l'artisanat par rapport à d'autres types de pratiques, tout en ouvrant la voie à des réflexions sur ses évolutions face aux défis contemporains tels que (4) l'automatisation et la numérisation.

Habiletés et communautés de pratique

- 7 L'artisanat, tel que le conçoit Sennett (2008), est à la fois une pratique et une éthique dont l'objectif principal est de « faire un bon travail pour lui-même ». Ici, la qualité prévaut sur l'efficacité, s'imposant comme la valeur cardinale (Kolko, 2011). Il s'agit d'un processus visant à créer un artefact fonctionnel, dans lequel conception et fabrication sont étroitement imbriquées (Kolko, citant Wroblewski, *ibid.*). Cette imbrication du design et de la fabrication en fait une pratique de cognition située, impliquant une « grande réflexivité sur les habiletés, les outils, les méthodes et les matériaux à mobiliser pour atteindre un objectif de faire les choses bien » (Schwalbe, 2010, p. 109, notre traduction).
- 8 L'artisanat repose ainsi sur le développement d'habiletés, fruit d'un apprentissage par essais et erreurs. Il exige un entraînement et un travail centré sur une tâche précise, parfois très simple, mais répétée continuellement (Kolko, *op. cit.*). Cette répétition permet à l'artisan·e d'acquérir une connaissance approfondie de son métier et de développer sa réflexivité. Comme l'illustre le travail de Portisch, cité par Ingold (2019), sur les artisanes en Mongolie, cette réflexivité et cette évaluation continues sont au cœur de la pratique, tant pour les novices que pour les artisanes expérimentées, et ce, à chaque étape du processus. L'acquisition et le développement des habiletés dans l'artisanat se fondent sur le principe de maître-apprenti·e et sur l'existence d'une communauté de pratique, définie comme le regroupement d'individus engagés dans une activité commune, quel que soit leur niveau d'expertise, et qui se caractérise par une structure évolutive où coexistent des membres expérimentés et des novices (Lave et Wenger, 1991). Aussi la pratique de l'artisanat et les valeurs qui la sous-tendent impliquent-elles que des personnes plus expérimentées partagent leur savoir avec d'autres personnes, nouvellement venues. Selon Klammer (2012, p. 7), puisque la valeur centrale de l'artisanat est la qualité, on désignera comme « maître » des personnes ayant la capacité de juger et de défendre la qualité dans la pratique, dont les habiletés et la créativité sont reconnues par les pairs et qui possèdent des qualités pédagogiques permettant la transmission du savoir. L'apprentissage, cependant, ne se limite pas à une simple observation passive. Il repose sur l'implication active de l'apprenti·e qui, au fil de son travail, interagit avec les artisan·es aguerri·es, observe leurs gestes et acquiert progressivement ses propres habiletés (Fyhn et Søraa, 2017). Cette relation favorise un apprentissage par la pratique, où la répétition des gestes s'accompagne d'une réflexion critique sur la qualité du travail accompli. Cette communauté de pratique permet alors non seulement l'acquisition progressive des compétences, mais renforce aussi l'ancrage des valeurs fondamentales de l'artisanat telles que la quête de qualité et le souci du travail bien fait (Kolko, *op. cit.*). La communauté de pratique joue ainsi un rôle essentiel dans la préservation, la transmission et la perpétuation de ces valeurs.

Care des choses et politique du faire

- 9 Cette quête de la qualité et ce souci du travail bien fait permettent de rattacher la pratique artisanale à une forme de *care*, entendu ici dans une perspective postmoderne. Le *care*, que l'on peut traduire par le fait de « s'occuper de », « faire attention », « prendre soin » ou encore « se soucier de » (Laugier et Molinier, 2009), a d'abord émergé dans la philosophie morale et les études féministes comme une manière de repenser l'éthique des relations interpersonnelles. Il s'agissait alors de mettre en valeur une éthique située, attentive aux vulnérabilités et aux interdépendances constitutives des relations humaines, tout en révélant la dimension genrée (c'est-à-dire presque exclusivement féminine) et souvent invisibilisée du travail du soin. Le concept a néanmoins progressivement migré vers d'autres disciplines, notamment les *science and technology studies* (STS), où il a été reformulé dans une perspective élargie aux non-humains.
- 10 Détaché de son ancrage exclusivement interpersonnel et genré, le *care* s'est vu appliqué aux objets, aux environnements et aux infrastructures, ouvrant la possibilité de penser des relations de soin qui ne concernent pas uniquement les humains, mais également les non-humains. Bruno Latour (2004), avec sa distinction entre *matters of fact* et *matters of concern*, a ouvert la voie en invitant à considérer les faits techniques et scientifiques comme des réalités toujours déjà enchâssées dans des préoccupations, des controverses et des attachements. Maria Puig de la Bellacasa (2011) a prolongé cette réflexion en introduisant la notion de *matters of care*, insistant sur le fait que ces préoccupations ne sont pas uniquement cognitives ou théoriques, mais impliquent également une dimension affective et éthique envers les non-humains.
- 11 Ce « soin des choses » comporte également une dimension politique (Laugier et Molinier, *op. cit.* ; Martin, Myers et Viseu, 2015). Prendre soin des choses, c'est non seulement garantir leur bon fonctionnement, mais aussi interroger les logiques qui président à leur production, à leur circulation, à leur persistance et à leur valorisation. L'artisanat se situe précisément dans cet espace. En réintroduisant du temps, de l'attention et de la minutie dans la production, il incarne une résistance discrète, mais structurante à l'hégémonie des logiques industrielles. C'est pourquoi plusieurs travaux soulignent la proximité structurelle entre artisanat et *care* (Kunneman, 2013 ; MacGill, 2019). Vasiliki Tsaknaki et Anna Vallgård (2023) rappellent ainsi que l'artisanat est fondamentalement « a matter of care ». Dans cette perspective, le soin et l'attention ne constituent pas des composantes secondaires, mais des principes constitutifs de la pratique artisanale. L'artisan·e ne se contente pas de produire un objet fonctionnel ; il ou elle exprime une forme d'engagement envers la matière travaillée.

L'artisanat comme rapport particulier à la matière

- 12 Certain·e·s auteur·rice·s, comme Klammer (*op. cit.*), défendent une conception rigoureuse de l'artisanat, centrée exclusivement sur le travail manuel et nécessitant des compétences acquises par l'entraînement dans le cadre d'une pratique soutenue. Dans cette perspective, un travail mobilisant des savoir-faire techniques et des habiletés spécifiques, mais où le geste manuel n'est pas central, ne saurait être considéré comme relevant, à proprement parler, de l'artisanat. Pour d'autres tels que Kuijpers (2019), l'essence de l'artisanat réside dans l'interaction dynamique entre l'artisan·e et la

matière. Selon ce dernier, « l'habileté, en tant que processus récursif et perceptif, crée des connaissances à partir des interactions matérielles » (p. 264, notre traduction). Ce n'est donc pas seulement la finalité de l'objet produit ou le geste qui importent, mais le processus même par lequel l'artisan·e développe sa compréhension du matériau à travers l'action. La matière façonne tout autant la pensée que la main : la pratique artisanale devient alors un dialogue entre l'artisan·e et le matériau.

- 13 L'exemple de l'âge du bronze illustre bien cette idée. Kuijpers (*ibid.*) montre que les propriétés spécifiques du bronze (sa malléabilité, sa durabilité et sa capacité à être fondu et remodelé) ont influencé les pratiques artisanales, facilitant l'émergence de techniques comme le design, le prototypage et la reproduction à grande échelle. Ces innovations ont profondément transformé les dynamiques sociales et économiques de l'époque, au point de donner naissance à une nouvelle ère.
- 14 L'artisan·e, dans ce processus, développe une connaissance intime du matériau par la compréhension de ses propriétés, de ses contraintes et de son histoire. Cette connaissance va au-delà d'une simple maîtrise technique, dans la mesure où elle intègre une dimension créative et réflexive qui permet non seulement de façonner la matière, mais aussi de l'altérer et de l'adapter à de nouveaux usages (Shorter, 2015, p. 4). L'artisan·e devient ainsi un·e médiateur·rice entre la matière brute et la forme achevée, inscrivant son geste dans une continuité historique et culturelle.

L'artisanat à l'ère des machines et de la numérisation

- 15 L'industrialisation et l'arrivée des machines dans le processus de fabrication ont provoqué plusieurs débats scientifiques quant à la notion d'artisanat en lien avec ce rapport à la matière (Fyhn et Søraa, *op. cit.* ; Kolko, *op. cit.* ; Schwalbe, *op. cit.*). Pour Ingold (*op. cit.*), la distinction fondamentale entre artisan·e et machiniste réside précisément dans leur relation à la matière. Cette réflexion s'appuie sur les travaux de Pye (1995), qui interroge la notion d'artisanat en la confrontant aux logiques de fabrication industrielle, notamment la production de masse. Selon lui, il est réducteur, voire inutile, d'opposer strictement le travail manuel au travail assisté par des machines ou des outils. Il propose plutôt de distinguer deux modes de fabrication : la fabrication à risque et la fabrication certaine.
- 16 La fabrication à risque implique une interaction continue entre l'artisan·e et la matière. Le résultat final dépend étroitement des compétences, de la dextérité et du jugement de l'artisan·e. Fyhn et Søraa illustrent cette idée en prenant l'exemple du menuisier : « Durant tout le processus, le menuisier doit rester alerte et pleinement engagé, en interaction directe avec la matière, car le résultat est en permanence à risque » (*op. cit.*, p. 75, notre traduction). L'instabilité propre à la création à risque crée ainsi un espace propice à l'expérimentation, à l'innovation et à la créativité. En revanche, dans le cas de la fabrication certaine, typique des environnements industriels, les machines sont configurées pour produire des résultats prévisibles et standardisés, sans considération pour le jugement ou la dextérité de l'opérateur·rice. Par exemple, si le même menuisier utilise une machine-outil pour découper les pièces d'une armoire en usine, il n'interagit plus directement avec la matière : il devient opérateur de la machine, perdant ainsi le lien sensible avec le matériau brut. Pye ne dénie pas pour autant les avantages de la fabrication certaine, particulièrement dans les contextes de production à grande échelle. Toutefois, il invite à réfléchir sur les pertes engendrées par l'effacement du

risque dans le processus. Schwalbe souligne également l'existence de ces pertes potentielles dans le cadre de la fabrication certaine, notamment « l'habitude de prendre soin dans le processus de travail, la dextérité et le jugement nécessaires à la confection manuelle, l'appréciation de la finesse du travail réalisé sans l'aide de la machine, les occasions d'innovation imprévue, la réflexion sur les limites de nos habiletés ainsi que l'esthétique de la diversité » (*op. cit.*, p. 110, notre traduction).

- 17 Traditionnellement, les recherches sur l'artisanat se sont concentrées sur la manipulation de matériaux physiques et sur la relation sensible entre l'artisan·e et la matière. Cependant, ces dernières années, un nouveau champ d'études s'est développé autour de l'artisanat numérique, qui interroge le rôle des outils numériques dans le processus créatif comme la production d'artefacts numériques (Treadaway, 2007 ; Zoran et Buechley, 2013 ; Grimshaw, 2017 ; Bernabei et Power, 2018). Cette approche s'intéresse à la manière dont les technologies numériques peuvent s'inscrire dans une logique artisanale en préservant des principes tels que la maîtrise technique et la sensibilité au détail lorsque la matière travaillée n'est plus physique.
- 18 Plusieurs chercheur·euses ont notamment élargi la notion d'artisanat aux pratiques numériques. Ces dernier·e·s soutiennent, par exemple, que la manipulation du clavier pour écrire du code peut être assimilée à un travail manuel, dans la mesure où elle mobilise des compétences techniques, de la créativité et un rapport direct à la « matière numérique » (Nachtigall, 2019). Dans le domaine du développement logiciel, cela s'est traduit par l'émergence du mouvement autour du *Manifeste pour l'artisanat du logiciel*. Né en réaction aux limites perçues des méthodes agiles, ce mouvement promeut une vision élargie de l'artisanat, centrée sur la qualité du code, la transparence, le sens des responsabilités ainsi que l'importance des dynamiques communautaires et du jugement des pairs (Sundelin *et al.*, 2021).
- 19 Finalement, une question persiste : les données sont-elles simplement au service de l'artisanat ou peuvent-elles devenir le matériau même de la création ? Cette question devient cruciale lorsqu'on se penche sur des formes d'artisanat où la matière (les données) se réduit à des éléments considérés comme abstraits — à savoir des 1 et des 0 —, comme c'est le cas dans la programmation ou la gestion de données (Shorter, *op. cit.*). Dans sa thèse consacrée aux liens entre artisanat et données dans le design industriel, Nachtigall (*op. cit.*) explore la notion d'artisanat des données et s'interroge sur leur matérialité. En s'appuyant sur les travaux de Dourish (2017), il souligne que les données, bien que souvent perçues comme immatérielles, reposent sur des infrastructures tangibles (code, systèmes, protocoles, machines) qui conditionnent la manière dont elles sont collectées, manipulées et visualisées. Cette matérialité confère aux données une forme de consistance qui les rend « façonnables » (Lambotte et Carillon, 2024). Nachtigall va au-delà de la question de la matérialité des données et s'intéresse davantage aux caractéristiques qui définissent l'artisanat des données en tant que pratique : l'apprentissage par la pratique, l'appartenance à une communauté, la maîtrise d'outils spécifiques et la place du jugement humain dans le processus de création. Selon lui, c'est par ces dimensions que l'on peut légitimement parler d'un artisanat des données, même lorsque la matière travaillée est numérique.
- 20 Notre approche s'inscrit ainsi dans la lignée des travaux de Nachtigall et propose de considérer l'artisanat des données comme un cadre d'analyse critique des processus de construction et d'exploitation des données. Dans ce contexte, nous considérons les données comme sociomatérielles, au sens où leur existence repose sur un ensemble

d'infrastructures qui participent de leur matérialité tangible (Bates, Lin et Goodale, 2016), mais aussi au sens de *matter* (Cooren, 2015), à savoir de leur poids dans les réalités sociales qu'elles constituent. Enfin, nous considérons l'artisanat des données comme une production « à risque », à savoir comme un espace propice à la réflexivité, à la créativité et à l'innovation et dont la persistance est remise en question par l'automatisation et la standardisation.

L'absence d'une perspective artisanale sur le travail des données

- 21 Avec l'intensification des processus de *datafication* (Mayer-Schönberger et Cukier, *op. cit.*), la question du travail des données (*data work*) s'est imposée comme un enjeu central dans la littérature scientifique. Dans la présente section, nous inscrivons notre proposition d'artisanat des données dans les études consacrées au travail des données. Cette exploration de la littérature sera l'occasion de montrer que, bien que certaines caractéristiques du travail artisanal soient parfois évoquées, aucun de ces travaux ne propose une conception théorique globale et cohérente permettant d'appréhender l'ensemble des dimensions spécifiques au travail artisanal des données.

Situer l'artisanat des données dans les études sur le travail des données

- 22 Le travail des données est défini par Bossen *et al.* comme toute activité « liée à la création, la collecte, la gestion, la conservation, l'analyse, l'interprétation et la communication des données » (*op. cit.*, p. 466, notre traduction). Englobant une vaste gamme d'activités et de pratiques (allant de l'encodage au nettoyage des données, en passant par leur standardisation, leur contextualisation ou la maintenance des bases), ce travail a longtemps été appréhendé par la littérature scientifique comme un « processus rationnel, guidé par les données, de "découverte" » (Rothschild *et al.*, *op. cit.*, p. 5, citant Muller *et al.*, *op. cit.*, notre traduction), qui relevait du ressort d'experts ou de spécialistes (Carter et Sholler, 2015). Cependant, comme le soulignent Rothschild *et al.* (*op. cit.*), la littérature en sciences des données et plus largement celle consacrée au travail des données ont amorcé, ces dernières années, deux décentrement successifs visant à dépasser ce cadrage restrictif. Notre proposition d'artisanat des données s'inscrit dans ces déplacements, tout en dialoguant avec les diagnostics critiques sur l'exploitation des formes invisibles de travail des données.
- 23 Le premier décentrement a consisté à déplacer l'attention des procédures techniques vers les pratiques situées et incarnées qui sous-tendent le travail des données. Ce déplacement a progressivement permis de déconstruire l'illusion d'objectivité et de rationalité attachée aux données, pour montrer que leur production est façonnée par des dynamiques sociales, cognitives et organisationnelles. Dans ce contexte, Koesten *et al.* (2021) identifient des schémas d'activités récurrents dans l'exploration et l'interprétation des données, révélant que ces tâches mobilisent des stratégies cognitives et des interactions sociales autant que des compétences techniques. De leur côté, D'Ignazio et Klein (2023) insistent sur la réflexivité nécessaire pour contextualiser et redonner sens aux données, tandis que Miceli, Schuessler et Yang (2020) mettent en évidence l'inscription de ces processus dans des rapports de pouvoir, où valeurs,

intérêts et priorités façonnent les résultats, en générant des frictions (Bates, Lin et Goodale, *op. cit.*).

- 24 Notre proposition d'artisanat des données s'inscrit pleinement dans ce décentrement. Le recours à l'artisanat permet, en effet, d'insister sur le caractère situé et incarné (Choroszewicz, 2022) de certaines pratiques en matière de production de données, dans la mesure où ces pratiques reposent sur des formes de jugement et d'expertise qui échappent aux représentations strictement techniques du travail des données. En effet, penser le travail des données en termes artisanaux revient à souligner que les données sont travaillées dans des contextes particuliers, mobilisent des expertises qui ne sont pas toujours codifiées et nécessitent des arbitrages continus qui associent savoirs expérimentiels et inscription dans des collectifs.
- 25 Le deuxième décentrement s'est opéré par l'élargissement de la focale au-delà des expertes pour inclure une pluralité d'acteur·rice·s intervenant dans la production et l'usage des données. Foster *et al.* (2018) proposent, par exemple, d'analyser le travail des données à travers trois niveaux (macro, méso et micro) et montrent que de nombreux profils, aux statuts et compétences variés, participent à la construction, à l'interprétation et à la circulation des données. En mettant en avant la figure de l'artisan·e et son inscription dans des communautés de pratique, l'approche artisanale prolonge ce déplacement en fournissant une manière de décrire concrètement comment cette pluralité s'organise dans certains contextes de production. Elle met en évidence la dimension collective du travail des données, fondée sur des communautés, des routines partagées et des dynamiques d'apprentissage qui assurent la cohésion et la continuité de ces pratiques.
- 26 Ce deuxième décentrement a également conduit la recherche à porter attention aux formes de travail périphériques (Rothschild *et al.*, *op. cit.*) qui soutiennent les infrastructures de données, c'est-à-dire aux activités des travailleuse·s de l'ombre « qui n'ont pas le rôle d'acteur, mais qui, par leur travail technique, permettent que la pièce se joue et que le public la voie » (Waquet, 2022, p. 8). C'est dans cette perspective que s'inscrit la littérature sur le *digital labor*, qui étudie les nouvelles formes de travail associées aux plateformes numériques et aux infrastructures de données, en mettant en évidence des dynamiques d'exploitation, de précarisation et de fragmentation (Casilli, *op. cit.*). Cette littérature couvre un spectre allant du micro-travail rémunéré sur des plateformes comme Amazon Mechanical Turk (où des tâches répétitives et peu qualifiées sont effectuées par des travailleuse·s dispersée·s et sous-payée·s) aux contributions non rémunérées des utilisateur·rice·s, dont les données sont captées et monétisées par les grandes entreprises du numérique. Gray et Suri (*op. cit.*) parlent ainsi de « travail fantôme » (*ghost work*) pour désigner ces tâches invisibles d'annotation, de nettoyage et d'organisation des données, tandis qu'Irani (*op. cit.*), à travers le « micro-travail » (*microwork*), insiste sur l'importance des modérateur·rice·s, tagueuse·s et classificateur·rice·s, indispensables au fonctionnement des plateformes, mais relégué·s à des positions marginales.
- 27 L'artisanat des données n'a pas pour vocation de minimiser la critique des rapports d'exploitation mise en évidence par le *digital labor*. Il s'agit plutôt d'un éclairage supplémentaire sur des pratiques concrètes qui, dans certains contextes de production et en fonction des finalités poursuivies, participent à conférer sens et valeur aux données. Cette approche permet de souligner que toutes les activités de production de données ne se réduisent pas aux seuls régimes d'exploitation et de précarisation,

notamment dans les contextes où le travail des données ne repose pas prioritairement sur l'extraction d'une plus-value économique immédiate, mais sur des finalités collectives, sociales ou épistémiques. Pour autant, ces formes de contribution demeurent insuffisamment reconnues ou valorisées et, partant, leur rôle dans le maintien des infrastructures de données reste largement invisibilisé (Denis, *op. cit.*).

Des « traits artisanaux » dispersés dans la littérature

- 28 Des « traits artisanaux » apparaissent de manière récurrente dans les travaux sur le travail des données, mais toujours de façon partielle, dispersée et sans donner lieu à une théorisation cohérente. Par traits artisanaux, nous entendons l'une des trois dimensions constitutives de l'artisanat identifiées plus haut, qu'il s'agisse d'une pratique orientée vers la qualité des données et inscrite dans des logiques de *care*, du développement d'habiletés collectives autour des données au sein de communautés de pratique ou encore d'un rapport « à risque » dans la production des données.
- 29 Les travaux sur les infrastructures de données, notamment, permettent d'entrevoir plusieurs de ces traits. Dans la lignée de Star (1999), l'objectif des *infrastructures studies* était, dès l'origine, d'attirer l'attention sur les « tâches concrètes » et les « ressources matérielles et symboliques » (Beltrame et Peerbaye, 2018, p. 179) qui assurent le maintien des infrastructures. Ces travaux mettent en évidence des pratiques qui vont de la réparation aux opérations de maintenance, en passant par des formes de bricolage situé (Harper, 1987) ou d'improvisation (Henke, 1999 ; Schubert, 2018). Appliquée aux infrastructures de données, cette perspective conduit Denis (*op. cit.*) à décrire ces activités comme un soin particulier porté à nos « infrastructures scripturales ». Dans ce contexte, Dagiral et Peerbaye (2022) mettent ainsi en évidence l'invisibilisation du travail effectué en amont de la collecte de données sur les maladies rares, soulignant la dépendance des infrastructures de données à des efforts méthodiques et minutieux réalisés par des acteur·rice·s souvent relégué·e·s à l'arrière-plan. De manière complémentaire, Stiefel (2018) prolonge cette réflexion en s'intéressant aux « agents opérationnels » dans la gestion des biobanques hospitalières. À travers une approche centrée sur le *care*, elle montre comment ces travailleuse·s assurent la continuité et la fiabilité des infrastructures de données biomédicales ainsi que les affects qui découlent de ce travail.
- 30 Ce lien entre travail des données et *care* a été exploré plus explicitement dans diverses publications (Zakharova et Jarke, 2024). Tandis que Jarke et Büchner proposent le concept de *data care arrangements* pour répondre à la question « pourquoi et comment les personnes dans les organisations se soucient des données » (2024, p. 702), Pinel, Prainsack et McKevitt montrent, dans leur étude de cas, que « les chercheurs établissent des relations à long terme avec les données qu'ils contribuent à produire et se sentent responsables de leur épanouissement et de leur développement ». Ils soulignent également que « ces pratiques de soin [...] contribuent à rendre les données précieuses pour l'institution » (2020, p. 175). Davies et Holmer (2024) étudient des dynamiques similaires dans le monde universitaire en mettant l'accent sur le fait que ces pratiques de *care* envers les données peuvent être lues comme des formes de résistance au discours dominant sur l'excellence universitaire. Enfin, Baker et Karasti (2018) approfondissent la dimension éthique et politique du *care* en mettant en lumière la manière dont le travail des données à l'échelle locale, dans le cadre du déploiement

de grandes infrastructures, est souvent relégué au rang de « chose négligée » et elles analysent les pratiques qui émergent en réponse à cette tension.

- 31 D'autres travaux mettent davantage l'accent sur les collectifs et les apprentissages dans le travail des données. Rothschild *et al.* (*op. cit.*), en mobilisant de manière explicite le concept de communauté de pratique, cherchent à dépasser une vision strictement occupationnelle des compétences pour mettre en évidence les dynamiques collectives, les valeurs partagées et les processus d'apprentissage qui structurent le travail des données. De telles dynamiques collectives se retrouvent dans plusieurs autres études (Cragin et Shankar, 2006 ; Benfeldt, Persson et Madsen, 2020). Néanmoins, si ces études traitent des dynamiques collectives dans le travail des données, elles restent cantonnées à l'analyse de la transmission des savoirs et de la socialisation professionnelle, sans les inscrire dans une conceptualisation plus large.
- 32 Pris ensemble, ces travaux montrent que les traits artisanaux sont donc présents dans la littérature, mais de manière dispersée, apparaissant en filigrane plutôt que comme les éléments d'un cadre unifié. C'est précisément cette dispersion que notre proposition vise à dépasser en proposant le concept d'artisanat des données.

L'artisanat des données au cœur de deux terrains ethnographiques

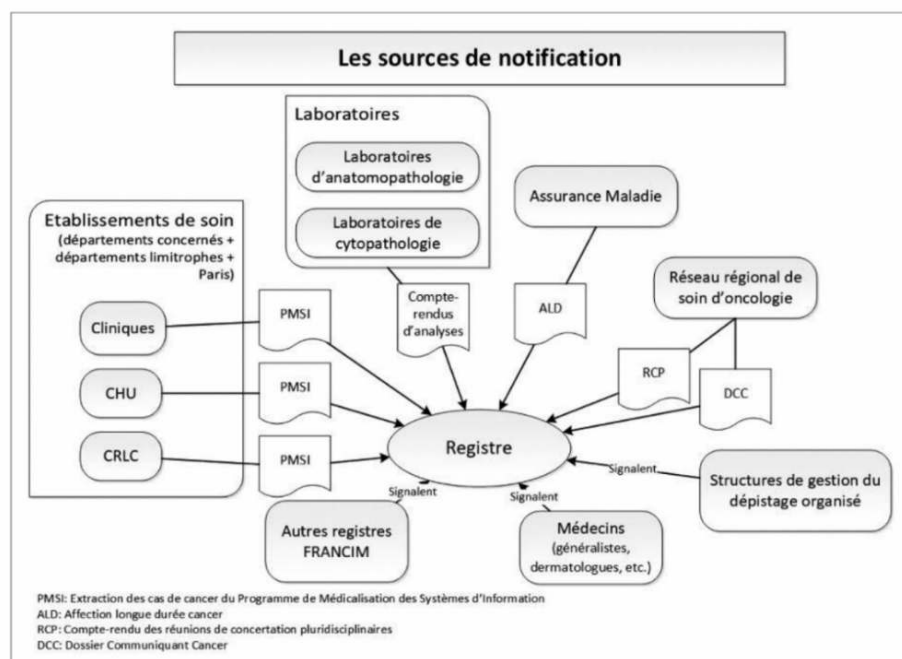
- 33 Si notre contribution est avant tout conceptuelle, elle ne se réduit pas à une abstraction détachée du terrain. Au contraire, nous entendons démontrer que l'artisanat des données s'ancre dans des réalités empiriques et permet, dans le même temps, d'éclairer des pratiques concrètes liées au travail de la donnée. Pour ce faire, nous nous appuyons sur des matériaux issus de deux terrains ethnographiques distincts, menés par l'un des auteurs. Ces données nous permettront d'examiner comment le travail des acteur·rice·s concerné·e·s s'apparente à un véritable artisanat des données, tout en mettant en exergue ses caractéristiques.
- 34 Nous nous appuyons d'abord sur un travail ethnographique (Martin-Scholz, Mayère et Lambotte, 2021) mené dans des organismes responsables de la production de RC dans deux départements français, qui ont pour mission de réaliser « une collecte continue et exhaustive de données nominatives concernant un ou plusieurs événements de santé [en l'espèce, le cancer] dans une population géographiquement définie, à des fins de recherche et de santé publique, par une équipe disposant des compétences appropriées¹ ». Le travail des RC peut être défini comme la transformation et l'assemblage de données initialement produites à des fins thérapeutiques ou économiques en données pouvant servir à des fins scientifiques et de santé publique. Ce travail se déroule dans le cadre d'un parcours des données, défini comme « le mouvement des données depuis leurs sites de production initiaux [vers le RC] dans lequel elles sont traitées, mobilisées et réutilisées » (Leonelli, 2020, p. 9).
- 35 Notre second terrain ethnographique (Renard *et al.*, 2024) est la communauté SO, un vaste réseau de forums en ligne où des communautés se rassemblent autour de sujets spécifiques. Fondée en 2008, la communauté SO est devenue une référence dans le domaine de la programmation. Sur cette plateforme, les utilisateur·rice·s posent des questions, apportent des réponses, votent sur la pertinence des contenus et peuvent les éditer.

- 36 Le point commun entre ces deux terrains de recherche est qu'ils regroupent des travailleur·euse·s de la donnée dont l'objectif est de produire des données de qualité. Dans un cas, ce travail est destiné à la recherche en sciences de la santé et à la lutte contre les cancers et, dans l'autre cas, il participe au domaine des sciences et des sciences appliquées. Ces deux collectifs organisés ont également en commun d'être bousculés par la « matérialisation des imaginaires de l'IA » (Bareis et Katzenbach, 2022), conduisant à l'émergence de controverses sur le rôle et les conséquences de l'IA sur le devenir des données produites et de l'infrastructure qui soutient celles-ci. Pour Pachidi *et al.* (2021), ces moments de controverse sont intéressants d'un point de vue analytique, car ils tendent à rendre visibles ces organisations ou communautés, qui se fondent habituellement dans les infrastructures dont elles assurent la maintenance.
- 37 C'est en travaillant à la compréhension de la transformation du travail de la donnée pour les RC, d'une part, et sur le mouvement de grève lancé par les modérateur·rice·s de la plateforme SO, d'autre part, que nous nous sommes rendu compte de la pratique située qui sous-tend la production des données. Cette prise de conscience repose sur plusieurs éléments : l'attention minutieuse que les acteur·rice·s portent à la donnée elle-même, les exigences de qualité qui lui sont associées, le savoir-faire nécessaire à sa production ainsi que les dynamiques collectives et les schèmes de valeurs qui structurent ces pratiques. C'est en observant ces processus et en interrogeant les activités des travailleur·euse·s de la donnée que l'image de l'artisan·e s'est progressivement imposée à nous. L'objectif des prochaines sections est donc d'aller au-delà d'une simple intuition et d'explorer ce que la conceptualisation de l'artisanat des données peut apporter à la compréhension du travail de la donnée.

Dans la fabrique des registres du cancer

- 38 En France, les RC sont répartis sur l'ensemble du territoire. Ces organisations de petite taille regroupent des enquêteur·rice·s dont l'objectif principal est d'identifier tous les cas de cancer pour une pathologie spécifique (par exemple, le cancer du sein) dans une région donnée. Une fois les cas répertoriés, l'équipe du RC doit rassembler les informations associées, qui sont souvent dispersées dans diverses bases de données (assurance maladie, hôpitaux, laboratoires, etc.) (voir figure 1).

Figure . Cartographie des sources de notification de la fabrique des données du registre du cancer



Source : Martin-Scholz, Mayère et Lambotte (op. cit.).

- 39 Ce processus implique plusieurs étapes cruciales. Les données doivent d'abord être collectées et centralisées, puis vérifiées et harmonisées au regard de la nomenclature internationale en vigueur, afin de garantir leur comparabilité et leur intégration dans le registre. La qualité des données est un principe fondamental de ce travail, reposant à la fois sur leur exactitude et leur exhaustivité. Ce processus de recontextualisation des données (Leonelli, 2016) requiert une expertise multidimensionnelle du fonctionnement des hôpitaux, de l'organisation du dossier patient, mais également des différences dans les usages des nomenclatures utilisées en milieu hospitalier et dans le cadre des registres.
- 40 Les RC, en tant qu'organisations, ne disposent pas d'un accès direct et systématique à toutes les bases de données hospitalières par l'intermédiaire des réseaux sécurisés (VPN). Par conséquent, les enquêteur·rice·s doivent se rendre sur place, consulter les dossiers physiques pour scanner et identifier les éléments pertinents. Bien que cette tâche puisse paraître simple, elle nécessite en réalité une connaissance approfondie de la structure des dossiers médicaux, des types d'examen à analyser et des éléments cliniques pertinents à extraire.
- 41 Lors de l'encodage, un autre défi émerge : les typologies utilisées pour caractériser le stade d'avancement d'un cancer peuvent varier entre les pratiques des oncologues, d'une part, et les standards internationaux adoptés par les RC, d'autre part. Cette discordance illustre la complexité du travail de recontextualisation, qui mobilise à la fois des connaissances médicales et une compréhension fine des normes administratives et scientifiques encadrant la production des données de santé.

[Codeuse RC] : Avec cela, j'ai mon [guide] de codage pour chaque topographie et quand je fais les poumons, je sors mon dossier sur les poumons et mes [guides] de codage [voir figure 2] parce que le codage est totalement différent d'un organe à

l'autre. J'ai donc là toutes mes feuilles d'enregistrement...

[Chercheur] : Vous annotez le guide au fur et à mesure en fonction de... ?

[Codeuse RC] : Au fur et à mesure. J'ai mis des notes sur [le recto du guide de codage] parce que ça a changé.

[Chercheur] : Je prends une photo de cette [figure 2] parce qu'il est intéressant de voir comment elle a évolué, ce qui a été ajouté.

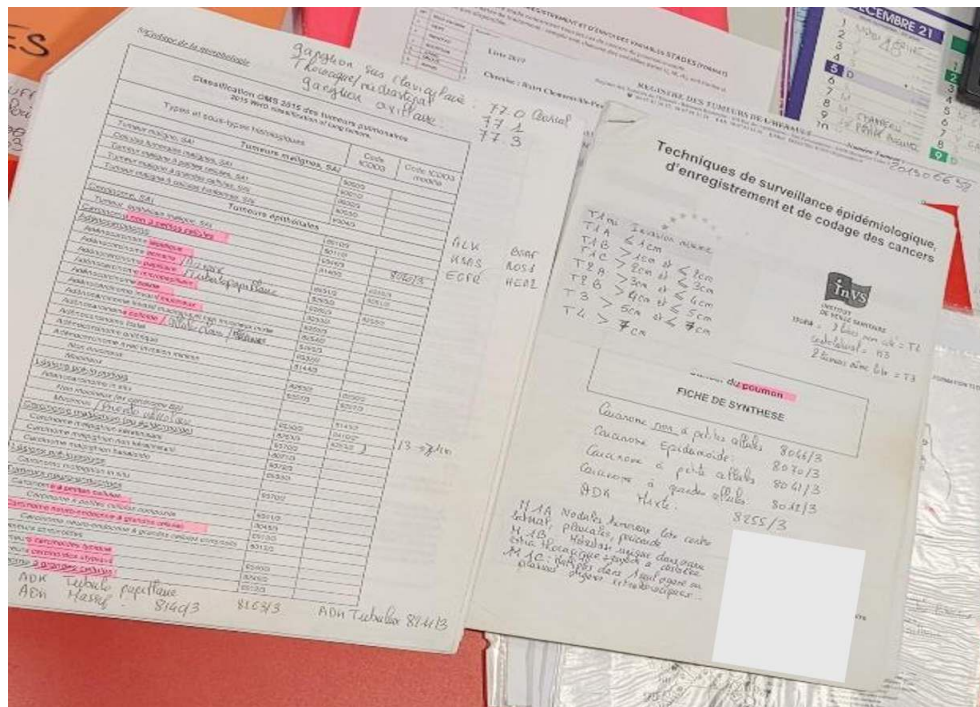
[Codeuse RC] : Certaines choses [dans la manière de coder et dans les informations enregistrées] ont changé, alors j'en ai pris note. On ne peut pas coder un estomac et un foie de la même manière. Donc, quand j'en fais un, j'en fais un. Je trie les organes.

[Chercheur] : Et chacun le fait à sa façon ?

[Codeuse RC] : Oh oui... Je trouve que c'est plus rapide de procéder ainsi.

- 42 Dans l'extrait ci-dessus, une codeuse explique la façon dont elle utilise un guide de codage pour traduire les informations du dossier médical dans le respect des nomenclatures définies à l'international par les RC. Les nombreuses annotations manuscrites sur son guide (figure 2) témoignent de l'apprentissage progressif et de la réflexivité qui accompagne la pratique répétée de l'encodage.

Figure . Guide de codage pour le cancer du poumon



- 43 Face à cette complexité, les enquêteur·rice·s s'appuient largement sur leurs collègues pour résoudre les incertitudes et affiner leurs pratiques. Notre ethnographie a ainsi révélé l'existence de dispositifs de formation continue et d'échanges de bonnes pratiques, notamment à travers des rotations entre le travail de collecte sur le terrain et l'encodage. Cette organisation favorise une meilleure compréhension mutuelle des différentes tâches et renforce l'entraide au sein de la communauté, comme en témoigne l'extrait suivant :

[Épidémiologiste responsable d'un RC] : 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 [...] En plus, nous, on essaie d'optimiser, [...] on va sur la source principale d'abord et si là, il n'y a pas assez, on va relancer les autres sources [...] 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 j'ai pas assez ».

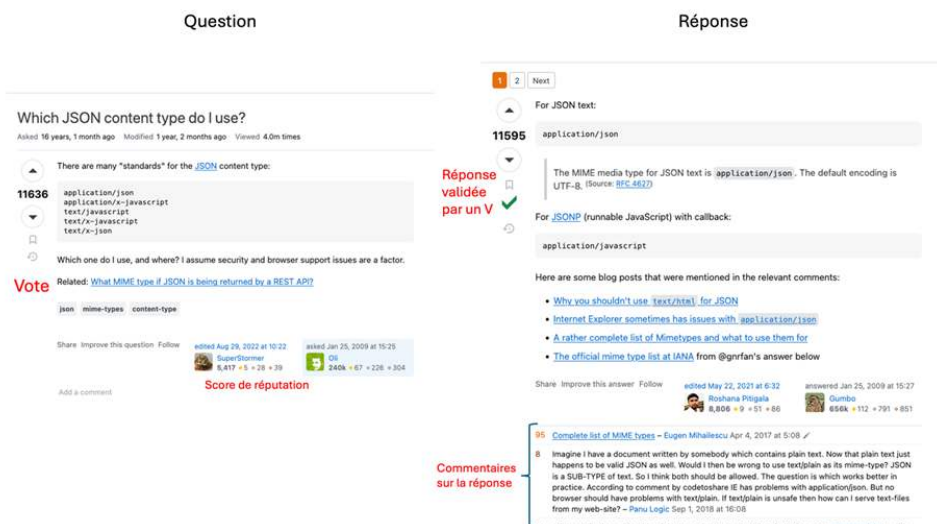
- 44 On constate ainsi que ce travail exige une réflexivité permanente, en particulier face aux avancées de la recherche sur le cancer. Par exemple, le cas d'étude révèle que l'épidémiologie impose une comparabilité historique des données afin d'analyser l'évolution des tendances sur un territoire donné. Cependant, l'amélioration des classifications et la découverte de nouvelles sous-catégories de cancers posent un dilemme : faut-il adapter les registres pour inclure ces distinctions plus fines, au risque de compromettre la continuité des données ? Ces décisions sont débattues et prises à l'échelle collective, au sein de la communauté des RC, représentée par l'association FRANCIM. Cette dernière joue un rôle-clé dans la standardisation des pratiques, en développant des outils logiciels et des guides d'encodage permettant d'harmoniser le travail des registres sur l'ensemble du territoire.
- 45 Le travail des enquêteur·rice·s met également en lumière une approche minutieuse et engagée vis-à-vis de la donnée. Ce soin se manifeste dans les efforts, parfois considérables, déployés pour récupérer des informations essentielles. Par exemple, lorsque l'accès aux données est bloqué par des laboratoires privés, les enquêteur·rice·s vont mettre en place des dispositifs bricolés pour récupérer le stade d'avancement d'une tumeur mentionné en annexe d'un dossier médical. Ce souci du détail dépasse la simple technicité de l'encodage : cet engagement contribue à garantir que les patient·e·s ayant moins de visibilité dans les grandes institutions hospitalières, par exemple, en raison de leur situation socioéconomique, soient correctement recensé·e·s dans les RC.
- [Enquêtrice RC] : Il est évident que si vous passez au big data, vous avez le nombre de cancers ! Mais ensuite, par stade, tout ça... on n'a pas la répartition... après, c'est un gros travail. Quand on parle aux médecins qui nous accueillent... quand ils voient le travail minutieux que nous faisons... parce que souvent, il faut chercher à plusieurs endroits. Je veux dire qu'une machine ne peut pas faire ça !
- 46 À travers l'image de la « machine » présent dans cet extrait, on constate également que les enquêteur·rice·s se positionnent par rapport aux imaginaires sociotechniques liés à l'IA et au *big data* (Bareis et Katzenbach, *op. cit.*). Pour ces dernier·e·s, ces imaginaires véhiculent une vision de la donnée comme une ressource brute, standardisée et immédiatement exploitable, occultant ainsi le contexte de sa production et la complexité de son traitement. Cette vision est problématique dans le cadre des RC, car elle tend à effacer l'importance et le rôle crucial des enquêteur·rice·s ; comme si le travail de sélection, de vérification et de contextualisation de la donnée n'avait pas de valeur. Or, l'automatisation ne peut remplacer l'expertise des enquêteur·rice·s, qui repose sur une compréhension fine des spécificités cliniques, administratives et épidémiologiques. Une approche purement automatisée risquerait de produire, au regard des enquêteur·rice·s, des données incomplètes ou biaisées, compromettant ainsi la fiabilité des analyses en santé publique et in fine des politiques de santé publique qu'elles alimentent. Pour ces artisan·e·s, il est donc essentiel de reconnaître et de préserver ce savoir-faire artisanal, qui constitue le socle de la qualité des données médicales utilisées pour la recherche et la lutte contre le cancer.

La communauté des modérateur·rice·s de Stack Overflow

Avec votre aide, nous travaillons ensemble à construire une bibliothèque de réponses détaillées et de haute qualité à chaque question sur la programmation².
Stack Overflow

- 47 La seconde communauté que nous avons choisi d'étudier est celle des modérateur·rice·s bénévoles chargé·e·s de l'édition et de la validation des contenus produits sur la plateforme collaborative SO. Fonctionnant selon un modèle de questions-réponses, SO permet à chacun de poser une question sur un sujet (généralement technique), à laquelle tout membre de la communauté peut répondre (voir figure 3). Comme l'indique la promesse affichée sur la page d'accueil et reprise dans l'extrait ci-dessus, l'ensemble des membres œuvre à garantir la qualité du contenu informationnel.
- 48 Le travail de cette communauté de pratique repose sur un dispositif sociotechnique complexe — c'est-à-dire une infrastructure scripturale (Denis, *op. cit.*) — fait de règles, d'outils, d'un mécanisme de gratification et de rôles construits au fil du temps par la communauté. La communauté est ainsi régulée par un code de conduite³ reprenant les normes de forme et de fond auxquelles doivent se conformer tous les contenus publiés. Par ailleurs, comme l'illustre la figure 3, les utilisateur·rice·s de la plateforme ont à leur disposition des outils permettant de sanctionner une question ou une réponse par un vote ou un signalement. Ce système de vote est central au fonctionnement de la plateforme : il attribue des points positifs ou négatifs aux contenus et génère des points de réputation pour leurs auteur·rice·s, leur octroyant au fil du temps des outils et des rôles supplémentaires, comme le fait de pouvoir éditer des publications.

Figure . Exemples de fonctionnalités de modération sur Stack Overflow



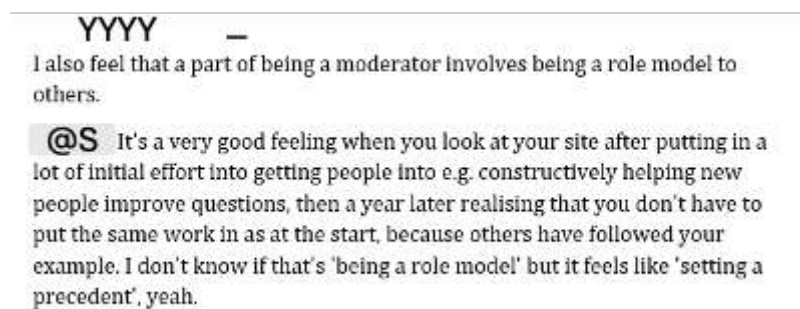
Légende : Stack Overflow, *Which JSON content type do I use?*, <https://stackoverflow.com/questions/477816/which-json-content-type-do-i-use>, page consultée le 11 avril 2025

- 49 La réflexivité joue un rôle central dans le travail de production et de modération. L'ensemble de la documentation nécessaire à ce travail ainsi que les discussions et décisions qui l'ont façonnée sont accessibles sur le site méta de la communauté⁴.

Récemment, à la suite de plusieurs différends avec l'entreprise qui héberge la plateforme, les modérateur·rice·s ont également mis en place un serveur Discord, où les membres discutent des règles et des outils, garantissant ainsi le bon fonctionnement de la communauté.

- 50 Ce qui frappe de prime abord lorsqu'on étudie cette communauté, c'est son organisation rigoureuse ainsi que les nombreux outils et procédures permettant à la fois de réguler l'afflux de contenus et le comportement des membres. Nous faisons ici l'hypothèse que ce dispositif sociotechnique complexe contribue en définitive à préserver le temps qualitatif de ces artisan·e·s bénévoles, qui peuvent ainsi se consacrer à l'accompagnement des utilisateur·rice·s (par l'entraide) et à l'amélioration du contenu informationnel (par l'édition de contenu).
- 51 Dans ce contexte, un mentorat par les pairs, assuré par les expert·e·s et les modérateur·rice·s, accompagne les personnes nouvellement membres. Comme l'illustre l'extrait ci-dessous (figure 4), issu d'une conversation sur le rôle des modérateur·rice·s, il existe une volonté affirmée d'agir de façon exemplaire pour les contributeur·rice·s ayant le moins d'expérience, au regard des standards de la communauté.

Figure 4. Extrait d'une conversation sur le rôle des modérateur·rice·s (Serveur Discord SO Moderators)⁵



- 52 L'attention portée aux contenus se manifeste à travers l'édition de contenu (les *handling flags* dans le jargon de la communauté). Cette exigence se traduit par une justification rigoureuse des décisions de modération, où chaque choix est argumenté et inscrit dans une logique de transparence et de fiabilité. L'édition des contenus témoigne ainsi d'un véritable travail sur les écrits, à l'image de celui des artisan·e·s façonnant la matière. Réalisé par les membres de la communauté, ce travail est au cœur de la préservation de la qualité et du soin apporté aux contenus présents sur la plateforme.

Figure . Question d'Artemis sur l'édition trop minutieuse d'un contenu⁶

What to do when a user targets your question history?

Asked 5 years, 3 months ago Modified 3 years, 9 months ago Viewed 674 times

▲ I'm not sure why, but recently, I had a single user (who only shares one tag with me) go through my question history, and make very questionable edits to posts that are both new (~1 month) and old (~8 months).

-5

▼ The changes are rather minute and unnecessary -

I am very confused, and am looking for someone to help explain that single line, and helping to map that one line into a much cleaner function.

I am seeking an explanation of that single line, and helping to map that one line into a much cleaner function.

Why is this user scanning through posts of mine from months ago and making minute grammar changes? This makes me feel targeted.

Note - I should add, I am all in favor of keeping high quality QA on Stack (look through my comment history -- I often write about this to new contributors). If this is justified, that is fine, but I would like to better understand *why* that is a necessary addition without an explanation to the OP on how to more accurately improve their questions in the future.

discussion edits

Share Follow

edited Dec 5, 2019 at 19:10

asked Dec 5, 2019 at 16:30

artemis 7,281 9 6

Figure . Réponse de Cody Gray défendant le travail d'édition des éditeur-ice-s⁷

▲ Yes, halfer does this. You aren't the first person to have noticed. :-)

40

▼ He's a prolific editor, devoting a significant chunk of his time to improving posts on Stack Overflow. Anecdotally, I don't think I've ever seen him submit an edit that I didn't think was an improvement over the original.

halfer is, in my experience, particularly adept at removing "noise": trimming and rephrasing the author's words into something that is easier for others to read, without losing the original meaning. It looks like that's what he's done here, for your questions.

Rest assured that being "targeted" in this way (having your posts edited) is not a bad thing. It shouldn't be taken to assume that you've done anything wrong—just that someone saw a way they could improve the site and took the opportunity.

Stack Overflow is collaboratively edited; this is one of the site's core principles. In this regard, and others, we share a lot of DNA with Wikipedia. Our primary goal, in fact, is to build a library of high-quality answers to the long tail of programming questions. Improving the presentation of individual questions is an important mechanism for achieving this goal. You might think that the edits are "minor"—why would anyone care if the wording of a question is a bit rambling?—but remember that Q&A aren't just here to help the original asker. They are maintained over the long term, where they help hundreds or thousands of people who have the same question and find the answers on our site via search engines.

I would like to better understand *why* that is a necessary addition at the expense of the OP.

I would like to better understand why you think that edits are "at the expense of the OP". Edits *help* the OP, and they help everyone else.

Share Follow

answered Dec 5, 2019 at 17:52

Cody Gray ♦ Mod 245k 84 728 770

13 that is exactly the type of response I was searching for. I appreciate the time you took to put together that response, and to attach and link some of the posts that you did. This is an excellent response. And, again for clarity, this was not an attack or knock -- *I just want to understand better so that I can one day make a bigger impact on the community.* Thanks! - artemis Dec 5, 2019 at 18:31

- 53 Dans la publication reproduite en figure 5, Artemis interroge le caractère parfois trop minutieux du travail d'un·e éditeur·rice et se demande si les efforts investis dans la

simple formulation d'une question présentent un réel intérêt. Dans la réponse qui s'ensuit (figure 6), Cody Gray illustre parfaitement la place centrale de l'édition dans la production de contenus de qualité. Il souligne ainsi que « ce que fait Halfer [l'éditeur-riche critiquée] consiste à enlever du bruit, à rendre les contenus plus faciles à lire pour autrui » et rappelle que l'édition collaborative est inscrite dans l'ADN de SO. C'est précisément ce travail qui, selon Cody Gray, permet de produire un contenu durable et de qualité.

- 54 SO répond ainsi pleinement aux critères d'une communauté de pratique : ses membres, tant les personnes les plus expérimentées que les nouvelles venues, partagent des objectifs communs, des méthodes, des valeurs, des règles et un vocabulaire spécifique. Les valeurs sous-jacentes de cette culture du « soin du contenu » s'ancrent dans l'éthique du *hacking*, qui promeut la création collaborative et décentralisée ainsi que la diffusion d'un savoir de qualité accessible à toutes et à tous (Turner, 2006).
- 55 En 2023, l'entreprise qui monétise les contenus produits par les bénévoles, en échange de la mise à disposition de l'infrastructure technique, a décidé d'interdire la suppression des contenus suspectés d'avoir été générés par l'IA. Cette décision a déclenché une grève des modérateur-riche-s de SO en réaction au manque de considération pour le travail accompli et parce que l'usage de l'IA générative les détourne de leur mission et de leur objectif.

Pour mes péchés, je suis le troisième plus grand éditeur de publications. L'édition semble désormais ancrée en moi, et je me demande si m'en défaire ne serait pas une épreuve. De plus, bien que nous, la communauté, voyions indéniablement la valeur de l'édition mutuelle, je ne suis pas certain que le retrait de ce travail préoccupe réellement l'entreprise⁸ (Halfer, notre traduction).

Je ne suis qu'un utilisateur « normal » qui utilise ses privilèges pour gérer le contenu au mieux de ses connaissances. Mais ces derniers jours, j'ai vraiment pris conscience de la situation et je me suis posé une question : Pourquoi est-ce que je continue à faire ça ? L'entreprise n'en a manifestement rien à faire, alors pourquoi devrais-je le faire ? C'est probablement la frustration qui parle, mais à ce stade, cela ne me dérangerait pas de laisser SE mourir dans une benne à ordures. À maintes reprises, les personnes qui se sont portées volontaires ici ont été déçues, alors pourquoi s'en préoccuper encore ? C'est pourquoi j'ai signé la lettre — Qbrute⁹ (notre traduction).

- 56 Le conflit autour des contenus générés par l'IA met en lumière plusieurs enjeux. Tout d'abord, comme illustré par le commentaire d'Halfer, cet épisode confirme que les modérateur-riche-s ont conscience de l'importance de leur travail dans le maintien d'un écosystème de savoir structuré, fiable et accessible. Dans ce contexte, les membres de la communauté établissent une nette distinction, en matière de qualité, entre les contenus produits et modérés par leurs soins et les contenus générés par l'IA. Par ailleurs, la grève apparaît comme une tentative pour faire prendre conscience à l'entreprise qui héberge la plateforme de l'importance de ce travail et des conséquences de sa décision quant à la qualité des contenus. Comme l'exprime Qbrute, la grève met ainsi en évidence le rôle des modérateur-riche-s et la nécessité de reconnaître leur contribution, mais aussi la nécessité de poursuivre le travail politique face à la pression grandissante de l'entreprise.
- 57 Enfin, la grève porte principalement sur la préservation du temps des modérateur-riche-s consacré aux pratiques artisanales des données. Or, en raison de la nouvelle politique de la plateforme, les modérateur-riche-s doivent désormais faire face à une surcharge de publications de faible qualité générées par l'IA, un phénomène qui détourne leur

énergie et leur expertise de leur mission première. En définitive, cette grève montre l'importance du rôle des modérateur·rice·s, non seulement aux yeux des utilisateur·rice·s, mais aussi de l'entreprise elle-même. La question de la qualité du contenu et de la reconnaissance du travail bénévole des artisan·e·s de la donnée se trouve ainsi au cœur des revendications.

Discussion et conclusion

- 58 En s'appuyant sur des données empiriques pour étayer notre proposition théorique, notre travail met en évidence plusieurs dimensions fondamentales de l'artisanat des données, qui seront discutées ici. Par ailleurs, nous proposons d'ouvrir la réflexion sur les tensions émergentes entre l'accumulation massive des données et l'exigence de qualité telle qu'on la trouve dans l'artisanat des données, en interrogeant la dimension politique de ce travail. Finalement et en guise de conclusion, nous reviendrons sur les limites et les perspectives qu'ouvre notre proposition théorique.

Retour sur la proposition d'artisanat des données

- 59 Bien que distincts dans leur finalité et leur organisation, les deux cas d'étude illustrent avec clarté que l'artisanat des données est une pratique sociale tout entière consacrée à la production de données de qualité. Les enquêteur·rice·s des RC, comme les modérateur·rice·s de SO, incarnent une forme d'artisanat de la donnée où l'expertise et le savoir-faire se construisent progressivement, à travers une interaction et une réflexivité constantes avec la matière première que représentent les données. L'artisanat des données, tel qu'on le trouve dans les deux cas d'étude, s'impose donc comme une pratique « à risque » (en opposition à la fabrication « certaine ») dans la mesure où ces pratiques artisanales opèrent dans un espace fait d'incertitudes, propice à l'expérimentation et à l'innovation, qui engage pleinement l'artisan·e dans son travail.
- 60 Ce souci constant de la qualité témoigne d'une attention minutieuse portée aux données, qui s'apparente à une forme de *care*, comme l'avaient déjà souligné plusieurs études (Stiefel, *op. cit.* ; Baker et Karasti, *op. cit.* ; Pinel, Prainsack et McKevitt, *op. cit.* ; Zakharova et Jarke, *op. cit.* ; Davies et Holmer, *op. cit.*). Dans cette perspective, le travail de soin s'incarne à travers des pratiques de maintenance et de réparation (Jackson, 2013 ; Denis et Pontille, 2022) des infrastructures de données et de leurs composants — en ce compris les données elles-mêmes (Baker et Karasti, *op. cit.*). Jackson met d'ailleurs l'accent sur cette filiation entre maintenance et soin porté aux choses en considérant que les pratiques qui en découlent réunissent « les mondes de l'action et du sens, reconnectant le travail nécessaire de maintenance avec les formes d'attachement qui, si souvent (mais de manière invisible, du moins pour les analystes), le soutiennent. Nous prenons soin parce que cela nous importe [*We care because we care*] » (*op. cit.*, p. 232, notre traduction).
- 61 Cette dimension s'avère d'autant plus prégnante que cette quête de qualité, couplée au soin des choses, revêt une portée éthique. En effet, les deux cas empiriques témoignent d'une conscience aiguë, chez les artisan·e·s, d'une forme d'éthique de la responsabilité dans la production et la gestion des données. Cette éthique de la responsabilité ne se limite pas à des considérations techniques, mais s'ancre dans un système de valeurs qui

guide leur travail au quotidien. Cette dimension est intimement liée à la finalité de leur activité, qu'il s'agisse de contribuer activement aux politiques de santé publique ou à l'utopie d'un Web ouvert, émancipateur et pourvoyeur de connaissances.

- 62 Cette responsabilité se manifeste de manière particulièrement marquée chez les enquêteur·rice·s et les représentant·e·s du CR, qui ne se contentent pas de produire des données fiables qui contribuent aux politiques de santé publique. Leur travail intègre aussi une réflexion critique sur la représentativité des minorités au sein des bases de données, afin d'éviter la reproduction d'inégalités systémiques. De manière analogue, dans la communauté SO, les artisan·e·s de la donnée s'investissent dans la démocratisation de l'apprentissage de la programmation. À travers le partage et la structuration de connaissances techniques sur la programmation, ces artisan·e·s bénévoles œuvrent pour un accès élargi aux savoirs, incarnant ainsi une vision émancipatrice de l'information et de la transmission du savoir.
- 63 Le concept d'artisanat des données permet toutefois de replacer cette éthique dans des dynamiques collectives. En effet, comme le montrent les deux cas d'étude, les artisan·e·s de la donnée évoluent également au sein de véritables communautés de pratique, rassemblant des membres aux niveaux d'expertise et d'expérience variables. Ces communautés forment le socle qui permet le développement des habiletés, elles favorisent les dynamiques d'apprentissage mutuel et de collaboration, où la production et la gestion des données sont des processus régulés et enrichis à travers l'échange avec d'autres membres. Ainsi, les artisan·e·s des données coconstruisent, partagent, valident et font circuler les connaissances, les méthodes, les valeurs selon une logique de coopération, garantissant non seulement la qualité des données produites, mais aussi la pérennisation (et, le cas échéant, l'évolution) des standards en vigueur dans la communauté.

Penser la dimension politique de l'artisanat des données

- 64 Dans les deux cas empiriques étudiés, la dimension politique de l'artisanat des données affleure à plusieurs reprises. Nous proposons de discuter cette dimension politique et son articulation avec les environnements datafiés à partir du concept de micropolitique. Développée par Deleuze et Guattari, la micropolitique repose sur l'idée que les centres de pouvoir sont multiples et que « tout est politique, mais toute politique est à la fois macropolitique et micropolitique » (1980, p. 260). Elle renvoie à la production de formes de pouvoir, de régulation et de gestion des relations sociales à une échelle locale, toujours en relation avec la macropolitique, qui s'exprime, quant à elle, à l'échelle des grandes structures institutionnelles et politiques.
- 65 La *datafication* (Mayer-Schönberger et Cukier, *op. cit.*), voire le dataïsme (van Dijck, 2014) relèvent de macropolitiques. Ils sont portés par des logiques institutionnelles, industrielles et gouvernementales qui privilégient l'accumulation massive, la standardisation et l'automatisation. L'artisanat des données peut alors être envisagé comme son actualisation à une échelle locale, où ces principes sont mis en œuvre, ajustés et contestés dans des pratiques concrètes de production et de gestion des données, orientées vers la qualité. Les artisan·e·s des données réintroduisent ainsi des formes de régulation échappant aux seules logiques industrielles, sans pour autant rejeter l'intérêt ou le potentiel des données et des technologies numériques dans leurs domaines respectifs. Leur travail consiste à négocier en permanence avec les

contraintes imposées par ces macropolitiques, tout en produisant, dans la pratique, des ajustements qui reflètent des préoccupations à la fois éthiques et pragmatiques.

- 66 Cette dimension politique s'exprime également dans la manière dont les artisan·es des données développent des savoirs situés et des pratiques de résistance discrètes face aux injonctions dominantes de la *datafication*. Dans un contexte où l'accumulation massive de données tend à être perçue comme une fin en soi, leur activité repose sur une exigence de qualité et de contextualisation, en rupture avec une vision extractiviste des données. Dans les deux cas étudiés, les artisan·es des données manifestent une vigilance particulière face aux risques que fait peser le déploiement de l'IA générative sur leurs pratiques et sur les valeurs qu'ils défendent. À l'instar du luddisme (Charitsis, Laamanen et Lehtiniemi, 2024), né au XIX^e siècle en réaction à la mécanisation du textile, l'IA est perçue comme une menace susceptible de conduire à la disparition des pratiques artisanales des données.
- 67 Cette tension est d'autant plus frappante que l'entraînement des IA génératives repose précisément sur des fondations solides : des données de qualité, sélectionnées et traitées avec soin. La dévalorisation apparente du travail de production et de gestion des données à l'ère de la *datafication* coïncide donc paradoxalement avec une dépendance accrue à ce travail pour garantir la fiabilité et la pertinence des résultats produits par les IA. Cette contradiction s'accroît encore avec l'essor des données synthétiques, c'est-à-dire artificiellement générées par l'IA, dont on sait désormais qu'elles alimentent des boucles autophages (ou « auto-consommatrices ») dégradant significativement la qualité des résultats produits par ces systèmes (Alemohammad *et al.*, 2023).

Limites et perspectives

- 68 Bien que certains traits artisanaux aient été relevés dans des recherches précédentes, notre analyse s'appuie seulement sur deux terrains empiriques et conserve donc, à ce stade, un caractère exploratoire. L'artisanat des données appelle dès lors à être étudié dans une pluralité de contextes empiriques. Notre proposition doit ainsi être entendue comme une invitation à approfondir l'investigation de ces pratiques par la recherche sur le travail artisanal des données. Au-delà de cette limitation empirique, plusieurs axes de réflexion méritent néanmoins d'être ouverts quant à la portée du concept.
- 69 Un premier enjeu concerne sa transférabilité. L'artisanat des données, tel qu'il se déploie dans les RC et sur SO, est-il susceptible de se manifester dans d'autres configurations de production ? La question se pose avec une acuité particulière dans les environnements caractérisés par une forte précarisation du travail, tels que ceux décrits par la littérature sur le *digital labor*. Cette dernière met en évidence des dynamiques d'exploitation, de standardisation et de fragmentation des tâches, qui semblent a priori incompatibles avec les caractéristiques de l'artisanat : autonomie relative, réflexivité, savoir-faire transmis et enrichi collectivement. Se pose dès lors la question de l'articulation entre ces deux registres : peut-on concevoir l'existence de zones de recoupement, où des micro-travailleurs élaborent, malgré tout, des pratiques artisanales (par exemple au travers de communautés informelles de soutien et d'entraide) ? Ou convient-il, au contraire, de penser l'artisanat comme un régime de travail radicalement distinct, réservé à des contextes institutionnels ou communautaires particuliers ?

- 70 Dans cette perspective, la finalité qui sous-tend l'artisanat des données apparaît également déterminante. Dans les RC, l'objectif de santé publique et la reconnaissance de l'expertise soutiennent une logique artisanale. Sur SO, la dimension communautaire, le bénévolat et l'idéologie open source favorisent la transmission et la valorisation des savoir-faire. Mais qu'en est-il dans les organisations privées, structurées par des impératifs commerciaux ? L'artisanat des données peut-il s'y épanouir alors que la logique de rentabilité tend à primer sur celle de la qualité ? Ou bien est-il condamné à être marginalisé, externalisé, voire subsumé par d'autres logiques organisationnelles ?
- 71 Enfin, la question de la reconnaissance sociale et institutionnelle se révèle décisive. L'artisanat des données demeure aujourd'hui largement invisible : il est faiblement valorisé dans les organisations, rarement reconnu dans les discours politiques et souvent occulté par la rhétorique de l'automatisation et les imaginaires de l'IA. Pourtant, comme nous l'avons montré, il constitue un maillon indispensable au fonctionnement des infrastructures numériques. Le saisir comme catégorie théorique permet non seulement d'en souligner la valeur propre, mais également d'ouvrir un horizon plus large lié à sa reconnaissance : comment rendre ce travail visible, lui accorder une reconnaissance sociale et institutionnelle, et mettre en place des dispositifs de protection et de soutien à cet artisanat à l'ère de la datafication ?

BIBLIOGRAPHIE

- ALEMOHAMMAD Sina, Josue CASCO-RODRIGUEZ, Lorenzo LUZI et Ahmed Imtiaz BARANIUK (2023), « Self-consuming generative models go mad », *arXiv preprint arXiv.2307.01850*, 4, p. 14, DOI : <https://doi.org/10.48550/arXiv.2307.01850>.
- BAKER Karen et Helena KARASTI (2018), « Data care and its politics: Designing for local collective data management as a neglected thing », *Actes du colloque 15th Participatory Design Conference: Full Papers-Volume 1*, p. 1-12.
- BARAIS Jascha et Christian KATZENBACH (2022), « Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics », *Science, Technology, & Human Values*, 47(5), p. 855-881.
- BARROWMAN Nick (2018), « Why data is never raw », *The New Atlantis*, 56, p. 129-135.
- BATES Jo, Yu-Wei LIN et Paula GOODALE (2016), « Data journeys: Capturing the socio-material constitution of data objects and flows », *Big Data & Society*, 3(2), DOI : <https://doi.org/10.1177/2053951716654502>.
- BELTRAME Tiziana et Ashveen PEERBAYE (2018), « Prendre soin des infrastructures. Introduction à la traduction de "L'ethnographie des infrastructures" de Susan Leigh Star », *Tracés. Revue de sciences humaines*, 35, p. 179-186.
- BENFELDT Olivia, John Stouby PERSSON et Sabine MADSEN (2020), « Data governance as a collective action problem », *Information Systems Frontiers*, 22, p. 299-313.

- BERNABEI Rina et Jacqueline POWER (2018), « Hybrid design: Combining craft and digital practice », *Craft Research*, 9(1), p. 119-134.
- BOCZKOWSKI Pablo J. et Kirsten A. FOOT (dir.), *Media technologies: Essays on communication, materiality, and society*, Cambridge, MIT Press, p. 221-239.
- BOSSEN Claus, Kathleen PINE, Federico CABITZA, Gunnar ELLINGSEN et Enrico Maria PIRAS (2019), « Data work in healthcare: An Introduction », *Health Informatics Journal*, 25(3), p. 465-474.
- CAI Li et Yangyong ZHU (2015), « The challenges of data quality and data quality assessment in the big data era », *Data Science Journal*, 14(2), p. 1-10.
- CARTER Daniel et Dan SHOLLER (2015), « Data science on the ground: Hype, criticism, and everyday work », *Journal of the Association for Information Science and Technology*, 67(10), p. 2309-2319.
- CASILLI Antonio (2015), « Digital labor : travail, technologies et conflictualités », dans Dominique CARDON et Antonio CASILLI (dir.), *Qu'est-ce que le Digital Labor ?*, Paris, Institut national de l'audiovisuel, p. 8-40.
- CHARITSIS Vassilis, Mikko LAAMANEN et Tuukka LEHTINIEMI (2024), « Towards algorithmic luddism: Class politics in data capitalism », *Information, Communication & Society*, 28, p. 1-18.
- CHOROSZEWICZ Marta (2022), « Emotional labour in the collaborative data practices of repurposing healthcare data and building data technologies », *Big Data & Society*, 9(1), DOI : <https://doi.org/10.1177/20539517221098413>.
- COOREN François (2015), « In medias res: Communication, existence, and materiality », *Communication Research and Practice*, 1(4), p. 307-321.
- CRAGIN Melissa et Kalpana SHANKAR (2006), « Scientific data collections and distributed collective practice », *Computer Supported Cooperative Work*, 15, p. 185-204.
- DAGIRAL Eric et Ashveen PEERBAYE (2022), « Les mains dans les bases de données. Connaître et faire reconnaître le travail invisible », *Revue d'anthropologie des connaissances*, 6(1), p. 191-216, DOI : <https://doi.org/10.3917/rac.015.0229>.
- DAVIES Sarah et Constantin HOLMER (2024), « Care, collaboration, and service in academic data work: Biocuration as “academia otherwise” », *Information, Communication & Society*, 27(4), p. 683-701.
- DE LA BELLACASA Maria Puig (2011), « Matters of care in technoscience: Assembling neglected things », *Social Studies of Science*, 41(1), p. 85-106.
- DELEUZE Gilles et Félix GUATTARI (1980), *Capitalisme et schizophrénie 2. Mille plateaux*, Paris, Éditions de Minuit.
- DENIS Jérôme (2018), *Le travail invisible des données. Éléments pour une sociologie des infrastructures scripturales*, Paris, Presses des Mines.
- DENIS Jérôme et David PONTILLE (2022), *Le soin des choses. Politiques de la maintenance*, Paris, La Découverte, coll. « Terrains philosophiques ».
- D'IGNAZIO Catherine et Lauren KLEIN (2023), *Data feminism*, Cambridge, MIT Press.
- DIMAS Geri Louise, Lauri GOLDKIND et Renata KONRAD (2023), « Big ideas, small data: Opportunities and challenges for data science and the social services sector », *Big Data & Society*, 10(1), p. 1-5.

- DOURISH Paul (2017), *The stuff of bits: An essay on the materialities of information*, Cambridge, MIT Press.
- FOSTER Jonathan, Julie MCLEOD, Jan NOLIN et Elke GREIFENEDER (2018), « Data work in context: Value, risks, and governance », *Journal of the Association for Information Science and Technology*, 69(12), p. 1414-1427.
- FYHN Håkon et Roger Andre SØRAA (2017), « Craftsmanship in the machine: Sustainability through new roles in building craft at the technologized building site », *Nordic Journal of Science and Technology Studies*, 5(2), p. 71-84.
- GITELMAN Lisa (2013), « *Raw Data* » is an oxymoron, Cambridge, MIT Press.
- GRAY Mary et Siddharth SURI (2019), *Ghost work. How to stop Silicon Valley from building a new global underclass*, Boston, Mariner Books.
- GRIMSHAW David (2017), « Crafting the Digital: Developing expression and materiality within digital design and manufacture », *The Design Journal*, 20(1), DOI : <https://doi.org/10.1080/14606925.2017.1352878>.
- HARPER Douglas (1987), *Working knowledge: Skill and community in a small shop*, Chicago, University of Chicago Press.
- HENKE Christopher (1999), « The mechanics of workplace order: Toward a sociology of repair », *Berkeley Journal of Sociology*, 44, p. 55-81.
- HOLMLID (dir.), *Nordes 2023: This space intentionally left blank*, 12-14 June, Linköping University, Norrköping, Sweden, DOI : <https://doi.org/10.21606/nordes.2023.47>.
- INGOLD Tim (2019), « Of work and words: Craft as a way of telling », *European Journal of Creative Practices in Cities and Landscapes*, 2(2), p. 5-17.
- IRANI Lilly (2015), « The cultural work of microwork », *New Media & Society*, 17(5), p. 720-739.
- JACKSON Steven (2014), « Rethinking Repair », dans Tarleton GILLEPSIE, Pablo J. BOCZKOWSKI, & Kirsten A. FOOT (Dir.), *Media technologies: Essays on communication, materiality, and society*, MIT Press Boston, p. 221-239.
- JARKE Juliane et Stefanie BÜCHNER (2024), « Who cares about data? Data care arrangements in everyday organisational practice », *Information, Communication & Society*, 27(4), p. 702-718.
- KLAMER Arjo (2012), « Crafting Culture: The importance of craftsmanship for the world of the arts and the economy at large », Actes du colloque *ACEI conference*, Kyoto, <http://www.klamer.nl/wp-content/uploads/2017/08/crafting.pdf>, page consultée le 9 octobre 2025.
- KOESTEN Laura, Kathleen GREGORY, Paul GROTH et Elena SIMPERL (2021), « Talking datasets – understanding data sensemaking behaviours », *International Journal of Human-Computer Studies*, 146, DOI : <https://doi.org/10.1016/j.ijhcs.2020.102562>.
- KOLKO Jon (2011), « Craftsmanship », *Interactions*, 18(6), p. 78-81.
- KUIJPERS Maikel H. G. (2019), « Material is the mother of innovation », dans Anna MIGNOSA et Priyatej KOTIPALLI (Dir.), *A Cultural Economic Analysis of Craft*, Cham, Springer International Publishing, p. 257-270.
- KUNNEMAN Harry (dir.) (2013), *Good work: The Ethics Of Craftsmanship*, Amsterdam, Uitgeverij SWP.

- LAMBOTTE François et Kévin CARILLON (2024), « Questionner les modes d'existence des traces numériques de nos échanges en organisation », *Approches théoriques en information-communication (ATIC)*, 8(1), p. 35-54.
- LATOUR Bruno (2004), « Why has critique run out of steam? From matters of fact to matters of concern », *Critical Inquiry*, 30(2), p. 225-248.
- LAUGIER Sandra et Pascale MOLINIER (2009), « Politiques du care », *Multitudes*, 37-38(2), p. 74-75.
- LAVE Jean et Etienne WENGER (1991), *Situated learning: Legitimate peripheral participation*, Cambridge, Cambridge University Press.
- LEONELLI Sabina (2016), « The philosophy of data », dans Luciano FLORIDI (dir.), *The Routledge handbook of philosophy of information*, New York, Routledge, p. 191-202.
- LEONELLI Sabina (2020), « Learning from Data Journeys », dans Sabina LEONELLI et Niccolò TEMPINI (dir.), *Data journeys in the sciences*, Berlin, Springer, p. 1-24.
- MACGILL Belinda (2019), « Craft, relational aesthetics and ethics of care », *Art/Research International: A Transdisciplinary Journal*, 4(1), p. 406-419.
- MARTIN Aryn, Natasha MYERS et Ana VISEU (2015), « The politics of care in technoscience », *Social Studies of Science*, 45(5), p. 625-641.
- MARTIN-SCHOLZ Anja, Anne MAYÈRE et François LAMBOTTE (2021), « Rendre compte du travail d'instauration et d'assemblage des données en santé : le cas de Registres des cancers en France », *Communication et organisation. Revue scientifique francophone en communication organisationnelle*, 59, p. 215-230.
- MASSANARI Adrienne (dir.), *Critical cyberculture studies*, New York, New York University Press, p. 255-269.
- MAYER-SCHÖNBERGER Viktor et Kenneth CUKIER (2013), *Big data: A revolution that will transform how we live, work, and think*, Boston, Houghton Mifflin Harcourt.
- MICELI Milagros, Martin SCHUESSLER et Tianling YANG (2020), « Between subjectivity and imposition: Power dynamics in data annotation for computer vision », *Actes du colloque ACM on Human-Computer Interaction*, 4(CSCW2), p. 1-25.
- MIGNOSA Anna et Priyatej KOTIPALLI (dir.), *A cultural economic analysis of craft*, Berlin, Springer International Publishing, p. 257-270.
- MULLER Michael, Ingrid LANGE, Dakuo WANG, David PIORKOWSKI, Jason TSAY, Vera LIAO, Casey DUGAN et Thomas ERICKSON (2019), « How data science workers work with data: Discovery, capture, curation, design, creation », *Actes du colloque CHI conference 2019 on human factors in computing systems*, New York, Association for Computing Machinery, p. 1-15.
- NACHTIGALL Troy Robert (2019), *Materializing data: Craftsmanship and technology for ultra-personalization*, thèse de doctorat sous la direction de Ron WAKKARY, Eindhoven, Technische Universiteit Eindhoven.
- PACHIDI Stella, Hans BERENDS, Samer FARAJ et Marleen HUYSMAN (2021), « Make way for the algorithms: Symbolic actions and change in a regime of knowing », *Organization Science*, 32(1), p. 18-41.
- PINEL Clémence, Barbara PRAINSACK et Christopher MCKEVITT (2020), « Caring for data: Value creation in a data-intensive research laboratory », *Social Studies of Science*, 50(2), p. 175-197.
- PYE Davide (1995), *The nature and art of workmanship*, Londres, Herbert Press.

- RENARD Damien, Salla-Maria LAAKSONEN, Sophie DEL FA, François LAMBOTTE et Jukka HUHTAMÄKI (2024), « Resistance(s) in platformed organizing », *40th EGOS Colloquium*, Milan, 4-6 juillet 2024, <https://dial.uclouvain.be/pr/boreal/object/boreal:292711>, page consultée le 12 septembre 2025.
- ROTHSCHILD Annabel, Amanda MENG, Carl DISALVO, Britney JOHNSON, Ben Rydal SHAPIRO et Betsy DISALVO (2022), « Interrogating data work as a community of practice », *Actes du colloque ACM on Human-Computer Interaction*, 6(CSCW2), p. 1-28.
- SCHUBERT Cornelius (2018), « Repair work as inquiry and improvisation: The curious case of medical practice », dans Philippe SORMANI, Alain BOVET and Ignaz STREBEL (dir.), *Repair work ethnographies: Revisiting breakdown, relocating materiality*, Singapore, Springer Nature Singapore, p. 31-60.
- SCHWALBE Michael (2010), « In search of craft », *Social Psychology Quarterly*, 73(2), p. 107-111.
- SENNETT Richard (2009), *The craftsman*, New Haven, Yale University Press.
- SHORTER Michael (2015), « The craft technologist », *Studies in Material Thinking*, 13, p. 1-19.
- STAR Susan Leigh (1999), « The ethnography of infrastructure », *American Behavioral Scientist*, 43(3), p. 377-391.
- STELMASZAK Marta (2022), « Inside a data science team: Data crafting in generating strategic value from analytics », *ECIS 2022 Research Papers*, 86, https://aisel.aisnet.org/ecis2022_rp/86, page consultée le 9 octobre 2025.
- STIEFEL Léa (2018), « Étudier le care en infrastructure. Les “petites mains” de la biobanque hospitalière », *Revue d'anthropologie des connaissances*, 12(3), p. 399-427.
- SUNDELIN Anders, Javier GONZALEZ-HUERTA, Krzysztof WNUK et Tony GORSCHKE (2021), « Towards an anatomy of software craftsmanship », *ACM Transactions on Software Engineering and Methodology*, 31(1), 6, p. 1-49.
- TREADAWAY Cathy (2007), « Digital crafting and crafting the digital », *The Design Journal*, 10(2), p. 35-48.
- TSAKNAKI Vasiliki et Anna VALLGÅRDA (2023), « Craft as a matter of care to inspire the design of computational things », *Actes du colloque Nordes 2023: This Space Intentionally Left Blank*, 12-14 juin, Linköping University, Norrköping, Sweden.
- TURNER Fred (2022) « Chapter 22. How digital technology found utopian ideology: Lessons from the first Hackers' Conference », dans David SILVER et Adrienne MASSANARI (dir.), *Critical Cyberculture Studies*, New York, University Press, p. 255-269.
- VAN DIJCK José (2014), « Datafication, dataism and dataveillance: big data between scientific paradigm and ideology », *Surveillance and Society*, 12, p. 197-208.
- WAQUET Françoise (2022), *Dans les coulisses de la science : techniciens, petites mains et autres travailleurs invisibles*, Paris, CNRS Éditions.
- ZAKHAROVA Irina et Juliane JARKE (2024), « Care-ful data studies: Or, what do we see, when we look at datafied societies through the lens of care? », *Information, Communication & Society*, 27(4), p. 651-664.
- ZORAN Amit et Leah BUECHLEY (2013), « Hybrid reassemblage: An exploration of craft, digital fabrication and artifact uniqueness », *Leonardo*, 46(1), p. 4-10.

NOTES

1. <https://www.becquerel.fr/le-patient/les-registres-des-cancers/>, page consultée le 3 mars 2025.

2. <https://stackoverflow.com/tour>, page consultée le 10 mars 2025. Notre traduction.

3. Stack Overflow, *Code of Conduct*, <https://stackoverflow.com/conduct>, page consultée le 11 avril 2025.

4. Par exemple, Stack Overflow, *Top Questions*, <https://meta.stackoverflow.com/>, page consultée le 11 avril 2025.

5. « yyyy — 10/07/2020 15:06 - Je pense également qu'une partie du rôle de modérateur consiste à être un modèle pour les autres.

@S C'est très gratifiant de voir le résultat de tous les efforts que vous avez fournis au début pour inciter les gens à aider de manière constructive les nouveaux venus à améliorer leurs questions, puis de constater un an plus tard que vous n'avez plus besoin de faire autant d'efforts qu'au début, car d'autres ont suivi votre exemple. Je ne sais pas si c'est "être un modèle", mais cela ressemble à "créer un précédent", oui » (Notre traduction).

6. « Je ne sais pas trop pourquoi, mais récemment, un utilisateur (qui ne partage qu'un seul tag avec moi) a parcouru l'historique de mes questions et a apporté des modifications très discutables à des publications récentes (environ 1 mois) et anciennes (environ 8 mois).

Les modifications sont plutôt minimales et inutiles. Pourquoi cet utilisateur parcourt-il mes publications datant de plusieurs mois et apporte-t-il des modifications grammaticales minimales ? Je me sens visé.

Remarque — je tiens à préciser que je suis tout à fait favorable au maintien d'une qualité élevée sur Stack (consultez l'historique de mes commentaires, j'aborde souvent ce sujet avec les nouveaux contributeurs). Si cela se justifie, très bien, mais j'aimerais mieux comprendre pourquoi cette modification est nécessaire sans qu'aucune explication ne soit donnée à l'auteur du message initial sur la manière d'améliorer plus précisément ses questions à l'avenir. » Notre traduction.

7. « Oui, halfer fait cela. Vous n'êtes pas le premier à l'avoir remarqué. :-)

C'est un éditeur prolifique qui consacre une grande partie de son temps à améliorer les publications sur Stack Overflow. Pour l'anecdote, je ne pense pas l'avoir jamais vu soumettre une modification qui ne constituait pas une amélioration par rapport à l'original.

D'après mon expérience, halfer est particulièrement doué pour supprimer le "bruit" : il raccourcit et reformule les mots de l'auteur pour les rendre plus faciles à lire, sans perdre le sens original. Il semble que c'est ce qu'il a fait ici, pour vos questions. Soyez assuré qu'être "ciblé" de cette manière (voir vos publications modifiées) n'est pas une mauvaise chose. Il ne faut pas en conclure que vous avez fait quelque chose de mal, mais simplement que quelqu'un a vu un moyen d'améliorer le site et en a profité.

Stack Overflow est édité de manière collaborative ; c'est l'un des principes fondamentaux du site. À cet égard, et à d'autres, nous partageons beaucoup d'ADN avec Wikipédia. Notre objectif principal est en fait de constituer une bibliothèque de réponses de haute qualité à la longue liste de questions sur la programmation. L'amélioration de la présentation des questions individuelles est un mécanisme important pour atteindre cet objectif. Vous pensez peut-être que les modifications sont "mineures" (pourquoi quelqu'un se soucierait-il de la formulation d'une question un peu confuse ?), mais n'oubliez pas que les questions-réponses ne sont pas seulement là pour aider la personne qui a posé la question à l'origine. Ils sont conservés à long terme, où ils aident des centaines ou des milliers de personnes qui se posent la même question et trouvent les réponses sur notre site via les moteurs de recherche.

J'aimerais mieux comprendre pourquoi vous pensez que les modifications sont "au détriment de

l'auteur du message initial". Les modifications aident l'auteur du message initial, mais elles aident aussi tout le monde » (Notre traduction).

8. Stack Overflow, *How can we handle troublesome contributors during the moderation strike?*, <https://meta.stackoverflow.com/questions/425722/how-can-we-handle-troublesome-contributors-during-the-moderation-strike>, page consultée le 11 avril 2025.

9. Stack Overflow, *What has happened to lead moderators to consider striking?*, <https://meta.stackoverflow.com/questions/424979/>, page consultée le 11 avril 2025.

RÉSUMÉS

Cet article introduit et développe le concept d'artisanat des données pour qualifier un ensemble de pratiques, souvent invisibilisées, qui participent à la production des données à l'ère de la *datafication*. En contraste avec les discours dominants centrés sur la quantité et l'automatisation, l'artisanat des données met en lumière des formes de travail minutieux, situé et collectif, qui confèrent leur qualité et leur pertinence aux données. À partir de deux enquêtes ethnographiques, l'article montre que, dans certains contextes de production, le travail des données repose sur des savoir-faire, une éthique du soin (*care*) et des dynamiques communautaires comparables à celles que l'on trouve dans l'artisanat. Cette proposition théorique permet d'analyser le rôle des artisans des données et d'explorer les tensions entre massification des données, exigence de qualité et transformations liées à la *datafication* et à l'essor de l'intelligence artificielle (IA). L'artisanat des données apparaît ainsi comme une pratique à la fois technique, éthique et politique, qui vise à réaffirmer la dimension fondamentalement humaine du travail des données.

This article introduces and develops the concept of data craft to qualify a set of often invisible practices that contribute to data production in the age of datafication. In contrast to the prevailing narratives around quantity and automation, data craft highlights forms of meticulous, situated, and collective work that result in high-quality, relevant data. Drawing on two ethnographic surveys, the article shows that in certain production contexts, data work is based on know-how, ethics of care, and community dynamics comparable to those found in handicrafts. This theoretical position examines the role of data artisans and explores the tensions between data proliferation, expectations of quality, and transformations related to datafication and the development of artificial intelligence (AI). Data craft thus appears as a technical, ethical, and political practice, which aims to reaffirm the fundamentally human dimension of data work.

Este artículo incluye y desarrolla el concepto de artesanía de los datos para designar un conjunto de prácticas, a menudo invisibles, que intervienen en la producción de la información en la era de la *dataficación*. En contraste con los discursos dominantes centrados en la cantidad y automatización, la artesanía de datos pone en evidencia, formas de trabajo minucioso, localizado y colectivo, que confieren su calidad y pertenencia a los datos. Con base en dos encuestas etnográficas, el artículo resalta que en algunos contextos de producción, el trabajo de los datos se basa en el saber hacer, una ética de cuidado (*care*) y dinámicas comunitarias comparables a las que se encuentran en la artesanía. Esta propuesta teórica permite analizar el rol de los artesanos de los datos e identifica las tensiones entre la masificación de los datos, la exigencia de la calidad y las transformaciones relacionadas a la dataficación y expansión de la inteligencia artificial (IA).

Por lo tanto, la artesanía de datos aparece, de esta manera, como una práctica a la vez técnica, ética y política que tiene como objetivo reafirmar la dimensión fundamentalmente humana del trabajo de los datos.

INDEX

Mots-clés : artisanat des données, datafication, communauté de pratique, care, travail invisible

Palabras claves : artesanía de los datos, comunidad de práctica, care, dataficación, trabajo invisible

Keywords : data craft, community of practice, care, datafication, invisible work

AUTEURS

FRANÇOIS LAMBOTTE

François Lambotte est professeur en communication organisationnelle à l'UCLouvain (Belgique) et membre du laboratoire LASCO (Laboratoire d'analyse des systèmes de Communication des organisations). Ses recherches visent, entre autres, à mettre en lumière les enjeux humains et organisationnels qui sous-tendent les fabriques de données dans différents contextes scientifiques et techniques. Courriel : francois.lambotte@uclouvain.be

KEVIN CARILLON

Kevin Carillon est doctorant à l'UCLouvain (Belgique) en sciences de l'information et de la communication au sein du LASCO (Laboratoire d'analyse des systèmes de Communication des organisations). Ses recherches portent sur les processus de *datafication* et d'algorithmisation dans les médias de service public, avec une attention particulière portée à leurs implications organisationnelles. Courriel : kevin.carillon@uclouvain.be