Notes pour Anne science ouverte

Je peux témoigner de deux expériences contrastées de mise à disposition des données de la recherche.

Je commence par leurs points communs : les deux ont été réalisées par des équipes scientifiques qui ont vraiment travaillé « en collectif » ; les deux ont débuté sans aucun moyen, et ont collecté des financements en cours de route ; les deux sont longitudinales, c’est-à-dire que les mêmes personnes sont réinterrogées à plusieurs reprises, ce qui permet de suivre des processus ; les deux portent sur les relations, les réseaux personnels, les parcours de vie et les rapports au travail. Les deux engagent les mixed methods, par des itérations entre données qualitatives et quantitatives.

Leur différence majeure est que l’une, qui a débuté en 1995,le Panel de Caen, n’avait au départ aucune prétention à être partagée et diffusée. Elle n’a donc été absolument pas préparée à ce processus !!

Pourtant, le laboratoire du CNRS dans lequel j’ai été recrutée en 1993 et dans lequel j’ai débuté ce panel, s’appelait Laboratoire d’Analyse Secondaire et des Méthodes appliquées à la Sociologie -LASMAS (Institut du Longitudinal). L’analyse secondaire était l’expression de l’époque (il y a 30 ans !) pour désigner la réutilisation de données (exclusivement statistiques) par d’autres personnes que celles qui avaient mené l’enquête. Pourtant, dès les années 2000, j’ai partagé ces données « de gré à gré » avec des chercheurs, doctorants, stagiaires, soit pour une publication commune, soit pour leur utilisation à eux, sans moi, pour des master, thèses, ou publications. Le protocole d’enquête a aussi été répliqué au Québec et en Argentine. J’ai donc progressivement, de fait, ouvert ces données, en guidant activement mes partenaires, en présentiel, aux temps de la préhistoire du partage de données en particulier qualitatives.

La seconde enquête a démarré bien plus récemment au moment du 1er confinement, en mars 2020, a été nommée VICO (la vie en confinement) puis PanelVICO. C’était le moment où l’idée de la mise à disposition commençait à vraiment percer, en particulier quelques revues commençaient à proposer de mettre des annexes en ligne avec les données, et l’ANR demandait de se prononcer sur ce point dans la rédaction des projets. La « science ouverte » devient maintenant de plus en plus une injonction. Dans le cas de VICO, dès la conception du projet, puis du questionnaire et du codage des réponses dans des fichiers, il a été prévu de les mettre à disposition, tant les données quanti que les données quali (ces dernières avec un délai demandé par les chercheurs). Et cette mise à disposition a été réalisée dès la première année de contractualisation avec l’ANR.

Cette énorme différence (mise à disposition imprévue/prévue) se mesure à la quantité de travail, d’astuce et d’inventivité qu’il a fallu déployer pour déposer la première, le panel de Caen, comparé à la seconde, VICO. J’aurais renoncé à cette ouverture des données du Panel de Caen si Anne n’avait pas repéré un appel d’offres de BeQuali qui paraissait tout à fait adapté. Mais il a vraiment fallu son implication en tant que documentaliste (nommer, classer, ranger…), et celle de Guillaume Garcia (ingénieur données au CDSP) pour que nous parvenions à préparer les textes et explications nécessaires à la prise en main de ces données qualitatives et quantitatives « sans moi ». Ce processus sera valorisé d’ailleurs en tant que tel par des articles (un juste déposé, un autre datapaper pour la rentrée), mais aussi par un document d’ « enquête sur l’enquête, un dépôt d’archives pour « l’histoire des sciences », le tout dans un objectif pédagogique et méthodologique.

Une des conclusions importantes de ces expériences contrastées est qu’il est important de ne pas considérer les données comme du matériau « en soi », arrivé spontanément et « tout neuf », neutre, sur le web. Elles sont plutôt le fruit de combinaisons complexes et chronologiques de théories, de méthodes, de questions scientifiques, d’hypothèses, d’axes d’interprétations, mais aussi de rencontres, de financements obtenus (donc de politiques publiques ou privées), de priorités des laboratoires, de négociations au sein des équipes de la recherche… et de processus biographiques des chercheurs. Il convient donc de ne pas faire illusion en gommant cette part d’imprévisible, de combinatoire, la dimension processuelle des recherches, surtout (mais pas seulement) lorsqu’elles s’étendent sur le long terme.