

Regards – Focus

Les producteurs professionnels de données sur la biodiversité face aux « *big data* » en écologie

Isabelle Arpin*

Sociologie, Université Grenoble-Alpes, INRAE, UR LESSEM, Grenoble, France

Résumé – Ce texte explore les réactions des producteurs professionnels de données au découplage entre production et utilisation de données sur la biodiversité, qui constitue, dans ce domaine, un phénomène marquant de la période actuelle. Cinq types de réactions sont identifiés : la satisfaction des producteurs que leurs données soient reprises par des spécialistes de l'analyse des données, leur attachement à des protocoles critiqués, leur intérêt pour des données jugées anecdotiques, la rétention de données et l'expression de scrupules face à l'extension de la production de données. Ce travail exploratoire montre que celle-ci revêt de multiples dimensions et remplit plusieurs fonctions, qui peuvent être altérées par le découplage entre producteurs et utilisateurs de données. Il invite également à prêter attention aux questionnements éthiques de ces producteurs.

Mots-clés : biodiversité / données / production de données / utilisation de données / professionnels

Abstract – Professional biodiversity data producers facing “big data” in ecology. The decoupling between biodiversity data producers and users has been described as a major feature of the current data era. This text explores how professional biodiversity data producers react to this decoupling, by drawing on a set of qualitative studies carried out over the last two decades and long-time participation in the discussions of data collection protocols within the scientific councils of Alpine protected areas. It identifies five types of reactions: the producers' deep satisfaction that the data they produce interest people specialized in biodiversity treatment and analysis, their attachment to criticized protocols, their interest in data deemed as anecdotal, data withholding, and the expression of scruples concerning the extension of data production to a rising number of species, individuals, habitats and aspects of life. This exploratory work shows that the decoupling between data producers and data users leads to focus on the aptitude of data production to respond to very specific research questions, whereas the production of biodiversity data fulfils several functions, such as reinforcing the ties between people interested in the same species or the attachment to a population or a place. It emphasizes the need to attend to the multiple functions of data production, to listen carefully to the ethical questionings of data producers, and to further study the link between biodiversity conservation and direct implication in data production.

Keywords: biodiversity / data / data production / data use / professionals

Il est souvent question d'une «révolution des données» (Kitchin, 2014) qui prendrait notamment la forme d'un « déluge ». La production massive de données concerne de nombreux domaines de la connaissance, dont celui de la biodiversité, où l'on assiste à une multiplication du nombre d'observateurs, des moyens d'observation et des entités observées. L'insistance sur l'accumulation de données présentée comme une spécificité contemporaine est toutefois contestée ou en tout cas relativisée par les historiens des sciences. Ces derniers soulignent que

différentes ères dans la production de données se sont succédé depuis la Renaissance avec, à chaque fois, le sentiment d'une production de données très supérieure à la capacité à les traiter avec les moyens de l'époque (Strasser et Edwards, 2017). Plus que par une explosion des données, la période actuelle se caractériserait par un découplage entre la production et l'utilisation des données (Hackett *et al.*, 2008 ; Baker et Millerand, 2010 ; Strasser, 2011, p. 2 ; Strasser et Edwards, 2017, p. 342-343 ; Maris, 2018, p 177-178).

De fait, les données sur la nature ont pendant longtemps été majoritairement utilisées par ceux qui les

*Auteur correspondant : isabelle.arpin@inrae.fr

avaient produites et qui connaissaient par conséquent leur contexte de fabrication, leur signification et leurs limites. Bien que de nombreux chercheurs en sciences de la vie continuent de produire les données sur lesquelles ils travaillent, les données sur la biodiversité sont de manière croissante utilisées par des spécialistes, bioinformaticiens, biostatisticiens et modélisateurs¹, qui n'ont pas toujours participé à leur production. Ce découplage s'observe de manière particulièrement claire, par exemple, dans des centres spécialisés comme le Centre d'étude et de synthèse de la biodiversité (CESAB) en France et les National Centers for Ecological Analysis and Synthesis (NCEAS) aux États-Unis, aujourd'hui considérés comme une des voies d'avenir et d'entrée dans la *big science* de l'écologie (Hackett *et al.*, 2008). Des chercheurs y agrègent et y traitent des données de natures et d'origines diverses et travaillent à des échelles spatio-temporelles et/ou sur des communautés plus larges que celles qui ont longtemps occupé les chercheurs en sciences de la vie. De tels dispositifs promettent de construire une intelligence de la biodiversité inaccessible au producteur-utilisateur de données, en faisant émerger des questions de recherche nouvelles, selon un raisonnement inductif, ou en testant des hypothèses, selon un raisonnement hypothético-déductif.

J'explore ici les réactions des praticiens de la gestion et de la conservation de la biodiversité au découplage entre production et utilisation des données, en m'appuyant sur plusieurs enquêtes qualitatives que j'ai conduites auprès d'eux ces deux dernières décennies ainsi que sur ma longue implication dans les discussions au sein des conseils scientifiques de parcs nationaux alpins sur les protocoles de production de données². Je commencerai par montrer que la littérature s'est peu intéressée aux producteurs professionnels de données, avant de présenter et de discuter leurs réactions au découplage entre production et utilisation des données. Je conclurai sur la nécessité de mieux prendre en compte

¹ Les bioinformaticiens (ou plutôt les informaticiens de la biodiversité : Soberon et Peterson, 2004 ; Sarkar, 2009) créent des bases et des infrastructures de données dont ils cherchent à faire émerger de nouvelles questions de recherche et de nouvelles connaissances en adoptant une démarche inductive. Les biostatisticiens privilégient plutôt une démarche hypothético-déductive. Ils formulent des questions et des hypothèses de recherche, qu'ils testent à l'aide d'outils statistiques appliqués aux données disponibles. Les modélisateurs élaborent des modèles d'évolution passée et future, qui requièrent pour leur paramétrage un grand nombre de données. Ces approches ne sont pas nécessairement exclusives et peuvent être combinées au sein d'un même programme de recherche.

² Ce texte développe le contenu d'une présentation aux journées du département Sciences pour l'action et le développement (SAD) de l'INRA le 11 octobre 2018.

les multiples dimensions et fonctions de la production de données, au-delà de la fourniture d'éléments de réponse à une question de recherche.

L'oubli des producteurs professionnels de données

Le découplage entre production et analyse des données a fait l'objet de plusieurs études. Devictor et Bensaude-Vincent (2016) ont mis en évidence son effet sur les données elles-mêmes : la constitution de vastes bases de données, qui a permis l'invention de la notion de biodiversité globale et son inscription à l'agenda politique international, s'est faite au prix d'une transformation des relevés naturalistes en données partiellement détachées de leur contexte (voir aussi Fortier et Alphandéry, 2017). D'autres travaux ont porté sur le partage de données entre chercheurs et les frictions qu'elles occasionnent (Edwards *et al.*, 2011) ainsi que sur la réutilisation de données (Zimmerman, 2008). L'utilisation par les chercheurs de données produites par des non-chercheurs a elle aussi été étudiée mais surtout dans le cas des membres d'associations naturalistes (Ellis et Waterton, 2004 ; Lawrence et Turnhout, 2010 ; Turnhout et Boonman-Berson, 2011). Fortier et Alphandéry (2017) ont souligné les phénomènes de résistance suscités par la standardisation, la bancarisation et l'ouverture des données : les producteurs bénévoles de données craignent en particulier que la méconnaissance des conditions de production et la poursuite d'objectifs qu'ils ne partagent pas nécessairement génèrent des erreurs d'interprétation et un mésusage des données. D'autres travaux ont mis en évidence la complexité des réactions des bénévoles face au devenir de leurs données, entre satisfaction de contribuer à l'avancement de la connaissance et aux politiques de conservation, sentiment de perte de contrôle et frustration liée à un manque de reconnaissance du travail accompli (Lawrence et Turnhout, 2010).

Le cas des données produites par les salariés des institutions publiques ou parapubliques de gestion et de conservation de la biodiversité a beaucoup moins retenu l'attention. Arts *et al.* (2015) indiquent que l'émergence de la « conservation digitale » va transformer le travail des praticiens de la conservation, mais sans entrer dans le détail de ces transformations. Or, ces praticiens produisent une part majeure des données sur la biodiversité, en particulier en France, où les espaces protégés, les conservatoires botaniques nationaux et l'Office français de la biodiversité, en particulier, contribuent de manière essentielle à la connaissance des espèces animales et végétales et de leurs habitats. Dès leur création, ces institutions ont produit des données qui ont dans un premier temps intéressé surtout leurs propres agents, les associations naturalistes et de protection de la

nature, le public, en ce qui concerne les espèces emblématiques, et certains chercheurs. À mesure que les séries de données sont devenues plus longues, que les moyens de les analyser se sont accrus et que la biologie de la conservation, apparue au milieu des années 1980, est devenue une discipline académique à part entière (Meine *et al.*, 2006), les écologues ont manifesté un intérêt croissant pour ces données. De plus en plus d'entre eux s'en saisissent et interviennent auprès des institutions de gestion et de conservation de la biodiversité pour réorienter les protocoles de production de données de manière à les ajuster à leurs objectifs.

Des réactions contrastées

Avant de présenter les réactions des producteurs de données, il faut rappeler le caractère trompeur du terme. Même dites «brutes», les *data* ne sont pas simplement données ni même simplement collectées mais toujours fabriquées par des dispositifs plus ou moins complexes et inscrits dans des institutions singulières (Latour, 1993 ; Gitelman, 2013). Dans le cas des données sur la biodiversité, ce travail de fabrication passe par une «connaissance par corps» (Bourdieu, 1997) acquise au fil d'une longue fréquentation des animaux ou des plantes et de leurs habitats. Il implique une capacité à s'ajuster voire à s'identifier aux espèces suivies, pour pouvoir les repérer et anticiper leur localisation et leur comportement, ce qui nécessite un apprentissage et un investissement personnel dans la durée et comporte une forte dimension affective (Lorimer, 2008). En dépit de leur aspect austère³, il y a dans les données quelque chose, et parfois beaucoup, de celles et ceux qui les ont produites. Désignant un placard où se trouvaient toutes les fiches remplies au cours de sa longue carrière, un botaniste employé par une association gestionnaire d'espaces protégés me dit lors d'un entretien : «toute ma vie est là». On comprend alors que la reprise de données par d'autres ne laisse pas leurs producteurs indifférents.

Satisfaction et attente de légitimation

Les salariés des institutions considérées expriment souvent, en particulier en public, le fait que leur travail consiste à respecter scrupuleusement les protocoles de collecte des données et à contribuer à la constitution d'une connaissance générale de la biodiversité, indé-

pendamment de leurs intérêts personnels. L'utilisation de leurs données par d'autres est alors vue comme la marque de la qualité et de l'utilité du travail effectué et l'opportunité d'améliorer les activités de connaissance et de gestion ou de conservation et de gagner en légitimité. Elle va dans le sens d'une *evidence-based conservation* (Sutherland *et al.*, 2004), construite sur le modèle de l'*evidence-based medicine* et accordant une place fondamentale à la constitution de bases de données centralisées sur la conservation, qui mettrait les institutions et leurs agents au moins partiellement à l'abri des critiques. La satisfaction de voir leurs données intéresser des chercheurs en écologie et en biologie de la conservation est donc la première réaction des producteurs professionnels de données sur la biodiversité.

Mais certaines difficultés liées au découplage entre production et utilisation des données apparaissent en situation d'entretien ou lors des sorties de terrain. Elles suggèrent un positionnement complexe et ambivalent, où la volonté de contribuer à une entreprise collective de connaissance et de conservation de la biodiversité le dispute à d'autres objectifs et considérations.

Attachement à des protocoles critiqués

Une première difficulté concerne l'attachement à des protocoles critiqués par les biostatisticiens pour leur manque de rigueur et de fiabilité des données produites. Les modalités d'estimation des effectifs des populations d'ongulés de montagne (chamois et bouquetins) constituent ici un bon exemple. Ces populations étaient extrêmement réduites au moment de la création des premiers espaces protégés et l'augmentation de leurs effectifs a longtemps été considérée comme un indicateur de la réussite de ces espaces. Le suivi de ces populations a été et continue dans une large mesure à être au cœur du travail quotidien et des préoccupations des gardes-moniteurs des parcs nationaux alpins. Pendant longtemps, il a notamment consisté en des comptages estivaux, souvent qualifiés d'exhaustifs, regroupant la quasi-totalité du personnel de terrain et visant à dénombrer le plus possible d'individus. Les biostatisticiens ont critiqué cette technique parce que la part d'animaux qui échappe au regard des agents varie d'une année sur l'autre, sans que l'on puisse savoir dans quelle proportion (Loison *et al.*, 2006). Ils ont proposé une autre méthode qui consiste notamment à dénombrer de manière répétée, chaque année, les individus le long d'un parcours prédéfini. Les comptages exhaustifs, par ailleurs coûteux, ont été progressivement abandonnés au profit de la nouvelle méthode. Nombre d'agents regrettent cette évolution, parce que les anciens comptages fournissaient un nombre global, facile à appréhender et à communiquer. Ils constituaient de plus une des rares occasions de rassemblement de tout le

³ Les données se présentent généralement sous la forme de tableaux où figurent *a minima*, pour chacune d'elle, le nom de son producteur, l'entité concernée, le lieu et la date de production de la donnée, des éléments sur les modalités et éventuellement les conditions de production de la donnée.

personnel, réparti dans des secteurs géographiques distincts, et de mise à l'épreuve des capacités et des habiletés de chacun à se déplacer en montagne et à repérer les animaux. Dans certains cas, les chasseurs étaient invités à participer, ce qui permettait d'échanger et de tisser des liens autour d'une espèce d'intérêt commun. Les comptages « exhaustifs » étaient en bref un moment fort, fédérateur et festif de la vie des parcs. Leurs responsables des activités scientifiques ne s'y sont d'ailleurs pas trompés. Dans le parc des Écrins, par exemple, ils continuent d'inviter les chasseurs et organisent un apéritif à l'issue de la journée.

Symétriquement, des agents montrent un faible entrain pour des protocoles élaborés selon les standards de la biostatistique. Par exemple, les plans d'échantillonnage aléatoire impliquent de prospecter aussi des sites que les agents savent peu favorables aux espèces visées et qui vont à rebours de leur inclination à aller préférentiellement là où « il y a des choses à voir » ou dans des endroits qu'ils apprécient particulièrement.

Intérêt pour des données jugées anecdotiques

Une autre difficulté ou du moins une divergence entre producteurs et utilisateurs de données concerne leur intérêt différentiel pour certaines données. Je prendrai ici l'exemple du Programme Proies Prédateurs (PPP), un programme scientifique multipartenarial qui visait à étudier l'impact de la prédation par le loup sur les populations d'ongulés sauvages dans le Mercantour. Le PPP impliquait de trouver des cadavres d'ongulés rapidement après la mort des animaux pour pouvoir en identifier la cause. Un des moyens utilisés pour trouver ces cadavres était de suivre des loups équipés de colliers indiquant leur position à des intervalles de temps rapprochés. L'idée était que des positions voisines, se traduisant à l'écran par des nuages de points, pouvaient correspondre à des sites de prédation que des agents de terrain devaient prospecter pour tenter de trouver des cadavres d'ongulés. Le programme s'est heurté à des difficultés à la fois techniques, administratives et politiques et a été abandonné. Ce qui m'intéresse ici, c'est le différentiel d'intérêt accordé aux données relatives à la position des loups équipés de colliers GPS. Ces données ont passionné les agents du parc national du Mercantour chargés de suivre les loups. Ils les ont attendues avec impatience, les ont analysées de près et commentées d'abondance. Ils les ont considérées en éthologues. Elles ont également beaucoup intéressé les chasseurs et d'autres acteurs locaux. Elles ont en revanche été jugées « anecdotiques » par les chercheurs responsables du PPP, parce qu'elles étaient pour eux un moyen d'accéder aux cadavres d'ongulés plutôt qu'une fin en soi et parce qu'elles ne concernaient qu'un très petit nombre d'animaux.

Rétention de données

Un garde-moniteur du parc national de la Vanoise rencontré pendant ma thèse m'a dit regretter s'être impliqué dans les comptages de chamois, au motif qu'« on n'avait pas vu que ça servirait à les tuer ». Le chamois est une espèce chassée et les quotas attribués aux associations de chasse tiennent compte de l'estimation des effectifs de la population.

J'ai retrouvé des réactions similaires dans le cas du loup, avec une intensité accrue du fait qu'il s'agit d'une espèce protégée et très emblématique. Un réseau « grands prédateurs » a été constitué qui a été progressivement étendu à l'ensemble des départements concernés par la présence des loups et des lynx en France. Il est principalement composé de professionnels appartenant à diverses institutions de gestion et de conservation de la nature (au premier rang desquelles l'Office français de la biodiversité et les espaces protégés). Les membres du réseau, appelés correspondants, sont chargés de recueillir des indices de la présence des prédateurs, notamment des crottes. Ces indices alimentent un processus complexe combinant notamment analyses génétiques et modélisations qui aboutit à une estimation de la taille de la population et de son risque d'extinction en fonction du taux de prélèvement (Granjou et Mauz, 2009). Ces éléments interviennent alors dans la négociation entre le gouvernement et les différentes parties prenantes pour aboutir à un quota de loups pouvant être abattus chaque année sans déroger aux engagements internationaux de la France. Des agents se disent gênés que leur travail contribue à un processus de régulation de la population de loups contraire à leurs convictions, au point pour certains d'entre eux de se retirer du réseau ou de ne pas communiquer certaines données.

Scrupules face à l'extension de la production de données

La production de données sur la biodiversité a d'abord concerné un petit nombre d'espèces, les plus emblématiques, souvent de grande taille (*a priori* les plus faciles à observer et à suivre). La diversification et la miniaturisation des moyens de suivi ont permis l'extension de la production de données à des espèces plus petites et à des milieux plus difficiles d'accès ; des drones sont depuis quelques années de plus en plus mobilisés, y compris dans les espaces protégés. Les animaux eux-mêmes sont parfois enrôlés dans leur propre observation, celle de leurs congénères ou de leurs milieux de vie (voir par exemple Riopel, 2018). Le nombre d'individus suivis pour chacune des espèces a également augmenté. Dans le cas du PPP, une fraction substantielle des populations d'ongulés présentes sur le territoire d'étude a été équipée

de colliers émetteurs. Les aspects de la vie que l'on cherche à connaître se sont eux aussi diversifiés : il s'agissait essentiellement, au départ, de savoir combien il y avait d'individus d'une espèce et où ils se trouvaient. La production de données s'est étendue aux aspects comportementaux, sanitaires, génétiques, adaptatifs et environnementaux (paramètres biophysiques de l'environnement). Dans un programme de recherche sur la marmotte alpine commencé dans les années 1990 dans un espace protégé, le nombre d'individus et de paramètres suivis a ainsi régulièrement augmenté. Les chercheurs se sont récemment intéressés à l'adaptation des marmottes au changement climatique : ils ont introduit un capteur thermique dans la cavité abdominale de certains individus pour suivre l'évolution de leur température corporelle pendant la période d'hibernation. La production de données sur la biodiversité concerne ainsi un nombre croissant d'espèces, d'individus de chacune d'elles, de milieux et d'aspects de la vie (Maris, 2018, p. 169-187).

Les enquêtes menées suggèrent l'existence d'un décalage de point de vue sur cette évolution entre producteurs et utilisateurs de données. Les utilisateurs souhaitent dans l'ensemble collecter un maximum de données, sur le plus grand nombre possible d'individus et à propos du plus grand nombre possible d'aspects de la vie. Pour un chercheur participant au PPP, « le rêve, ce serait d'équiper tous les ongulés ». De leur côté, certains praticiens de la gestion et de la conservation de la biodiversité sont troublés par cette extension de la production de données. Dans le Mercantour, des agents du parc national se sont progressivement tenus à l'écart du PPP parce qu'ils estimaient qu'« il faut laisser les loups en paix ». Certains ont même renoncé à suivre les traces de loups en hiver pour éviter de les déranger, en dépit de leur intérêt passionné pour l'espèce. En Vanoise, le programme de recherche sur la marmotte alpine a suscité des questionnements et des critiques de la part d'agents de terrain, gênés que des animaux puissent subir des manipulations répétées et des interventions chirurgicales dans un espace protégé, même au nom de la science.

Discussion

Contrairement à celles des bénévoles, les réactions des praticiens de la gestion et de la conservation de la biodiversité au découplage entre production et utilisation de données ont été peu étudiées. Ce faible intérêt s'explique sans doute par le fait que les professionnels semblent plus faciles à discipliner que les bénévoles, la production de données faisant partie de leurs « missions » (voir Lawrence et Turnhout, 2010, p. 6). Leurs réactions témoignent pourtant des enjeux sociaux, politiques, affectifs et éthiques des activités liées à la production,

l'archivage, la circulation et l'utilisation des données. Elles apparaissent assez proches de celles des bénévoles. La scission avec les utilisateurs de données suscite en particulier la même ambivalence que chez ces derniers (Lawrence et Turnhout, 2010). Toutefois, la satisfaction liée au gain de légitimité procuré par la reprise des données par des spécialistes de leur traitement s'exprime peut-être plus franchement chez les professionnels, tandis que les formes de résistance sont, elles, plus discrètes. L'organisation du travail des agents des institutions de conservation et de gestion de la nature leur permet souvent de trouver des arrangements plutôt que de protester frontalement (ce qu'ils font aussi parfois), notamment en s'investissant préférentiellement dans les actions qui les intéressent le plus et en se tenant à l'écart des programmes qui ne satisfont pas leurs goûts ou choquent leurs convictions naturalistes.

Trois autres points me semblent particulièrement importants à discuter : les fonctions multiples de la production de données, la propriété des données et l'extension de leur production.

Pourquoi produit-on des données ?

Les réactions identifiées montrent que la production de données a plusieurs fonctions. Or, lorsqu'elle est découplée de l'utilisation, la capacité des données à répondre à une question de recherche devient généralement la préoccupation prioritaire. L'équation « un protocole = une question de recherche » tend ainsi à se diffuser dans les institutions de gestion et de conservation de la biodiversité. Il ne s'agit pas de remettre en cause cette approche mais de pointer, pour les prévenir, deux risques potentiels qu'elle comporte.

Premièrement, le risque d'une vision appauvrie de la production de données. On l'a vu avec le PPP, les données peuvent intéresser des acteurs qui se posent différents types de questions relevant de plusieurs disciplines, en l'occurrence l'éthologie et la biologie des populations. Par ailleurs, et surtout, produire des données sert aussi, souvent sinon toujours, à autre chose qu'à répondre à une ou plusieurs questions de recherche, de même qu'un repas pris en commun ne sert pas seulement à nourrir les convives. Par exemple, produire des données permet également d'établir ou de renforcer des liens entre des personnes qui s'intéressent aux mêmes entités de nature mais sont géographiquement dispersées, et de cultiver des attachements à des espèces, des populations, voire des individus et/ou des lieux.

Deuxièmement, le risque d'une réponse insuffisante aux réticences des producteurs à répondre favorablement aux demandes des utilisateurs de données. Trois grands types de réponses sont classiquement proposés dans la littérature : la formation aux bases de la bioinformatique

et de la biostatistique (voir par exemple [Soberon et Peterson, 2004](#), p. 696), la restitution des résultats des recherches des utilisateurs et la reconnaissance de l'importance du travail des producteurs. Entendons-nous bien : ces réponses sont absolument nécessaires. Mais elles ne sont pas suffisantes. Il faut aussi faire en sorte que l'ensemble des fonctions de la production de données continuent d'être assurées lorsque les activités de production et d'utilisation sont découplées. Sauf à envisager, à l'avenir, de confier la production de données sur la biodiversité à des machines, avec les problèmes qui en découlent ([Sandbrook, 2015](#)).

À qui appartiennent les données ?

[Strasser et Edwards \(2017, p. 343\)](#) observent que les données appartiennent de manière croissante à ceux qui les utilisent. Ce sont eux qui décident à quoi elles doivent servir et comment elles doivent être produites. La rétention de données par des producteurs professionnels peut être interprétée comme une réaction à cette forme d'appropriation des données par les utilisateurs. Elle traduit le fait que les producteurs de données s'estiment fondés à exercer une sorte de droit de suite sur ces dernières, précisément parce qu'ils les ont produites et parce qu'ils ont une connaissance précise de la façon dont elles l'ont été et donc de leur signification. Son ampleur est par essence difficile à cerner même si elle semble rare chez les professionnels. Elle peut intervenir lorsque les producteurs de données estiment que leur travail est insuffisamment reconnu ou lorsque les données sont utilisées à des fins qu'ils n'avaient pas anticipées et qu'ils déplorent.

Jusqu'à où produire des données ?

La production de données a des effets sur les entités de nature : il faut capturer les animaux pour les équiper et les suivre ; il faut instrumenter les habitats pour produire des données biophysiques, ce qui nécessite de parcourir les milieux, souvent avec des véhicules motorisés, terrestres ou aériens. Les relations avec ces entités de nature s'en trouvent elles aussi modifiées. Les animaux, les plantes et les écosystèmes deviennent bons à connaître et à gérer à travers la connaissance acquise.

Parce qu'ils sont au plus près de ces entités, les producteurs de données ont une conscience aiguë de ces effets. Mes observations suggèrent que cette proximité peut être à l'origine d'une forme de souci, de *care*, qui pourrait expliquer que les producteurs aient davantage de scrupules que les utilisateurs à étendre encore la production de données. Le découplage entre production et utilisation de données pourrait alors avoir des effets collatéraux imprévus sur la conservation de la biodiversité, qui demandent à être étudiés de près.

Conclusion

L'observation des réactions des producteurs professionnels de données sur la biodiversité au découplage entre production et utilisation des données incite en définitive à être attentif :

- aux différentes dimensions de la production de données ; ces dimensions sont certes scientifiques et techniques mais aussi politiques, sociales, affectives, éthiques ;
- à l'ensemble des fonctions qu'elle remplit. Y compris dans le milieu professionnel, elle engendre bien autre chose que des données en réponse à une question de recherche. Or, le découplage croissant entre production et analyse de données peut favoriser une approche restrictive et appauvrie de la production de données et peut altérer ses fonctions moins visibles ;
- aux questions d'ordre politique et éthique que posent les producteurs de données et à leurs scrupules. Ils contribuent clairement à l'extension de la production de données mais, au plus près des entités de nature suivies, ils sont aussi bien placés pour percevoir ses risques et ses enjeux, pour les populations et les individus visés ainsi que le rapport que nous entretenons avec eux et avec l'activité de connaissance.

Pour finir, je voudrais souligner, sans pouvoir développer ici, que les questions que je viens d'évoquer concernent aussi les chercheurs en sciences sociales. Nous utilisons en effet des données produites avec d'autres, nos « informateurs » et nos « informatrices », très souvent sans que ces autres aient leur mot à dire dans la manière de produire, d'archiver et d'analyser ces données. À qui appartiennent les données, pourquoi les produire, jusqu'à où le faire et quand s'arrêter, comment décide-t-on que certaines données sont intéressantes ou pas ? Toutes ces questions soulevées par les producteurs professionnels de données sur la biodiversité doivent aussi nous préoccuper quand nous menons des entretiens, observons les pratiques de nos informateurs et utilisons ensuite des données qui ont clairement été coproduites.

Références

- Arts K., van der Wal R., Adams W.M., 2015. Digital technology and the conservation of nature, *Ambio*, 44, Suppl. 4, S661-S673.
- Baker K.S., Millerand F., 2010. Infrastructuring ecology: challenges in achieving data sharing, in Parker J.N., Vermeulen N., Penders B. (Eds), *Collaboration in the new life sciences*, London, Ashgate.
- Bourdieu P., 1997. *Méditations pascaliennes*, Paris, Seuil.

- Devictor V., Bensaude-Vincent B., 2016. From ecological records to big data: the invention of global biodiversity, *History and Philosophy of the Life Sciences*, 38, 4, 13.
- Edwards P.N., Mayernik M.S., Batcheller A.L., Bowker G.C., Borgman C.L., 2011. Science friction: data, metadata, and collaboration, *Social Studies of Science*, 41, 5, 667-690.
- Ellis R., Waterton C., 2004. Environmental citizenship in the making: the participation of volunteer naturalists in UK biological recording and biodiversity policy, *Science and Public Policy*, 31, 2, 95-105.
- Fortier A., Alphanéry P., 2017. La maîtrise des données, un enjeu majeur pour les associations naturalistes à l'heure de la gouvernance de la biodiversité, *Revue française d'administration publique*, 163, 3, 587-598.
- Gitelman L., 2013. "Raw Data" is an oxymoron, Cambridge/London, The MIT Press.
- Granjou C., Mauz I., 2009. Quand l'identité de l'objet-frontière se construit chemin faisant. Le cas de l'estimation de l'effectif de la population de loups en France, *Revue d'anthropologie des connaissances*, 3, 1, 29-49.
- Hackett E.J., Parker J.N., Conz D., Rhoten D., Parker A., 2008. Ecology transformed: the National Center for Ecological Analysis and Synthesis and the changing patterns of ecological research, in Olson G.M., Zimmerman A.S., Bos N. (Eds), *Scientific collaboration on the Internet*, Cambridge/London, The MIT Press, 277-296.
- Kitchin R., 2014. *The data revolution. Big data, open data, data infrastructures and their consequences*, London, Sage.
- Latour B., 1993. Le topofil de Boa Vista ou la référence scientifique-montage photophilosophique, in Conein B., Thévenot L., Dodier N. (Eds), *Les objets dans l'action. De la maison au laboratoire*, Paris, Éditions de l'EHESS, 187-216.
- Lawrence A., Turnhout E., 2010. Personal meaning in the public sphere. The standardisation and rationalisation of biodiversity data in the UK and the Netherlands, *Journal of Rural Studies*, 26, 4, 353-360.
- Loison A., Appolinaire J., Jullien J.-M., Dubray D., 2006. How reliable are total counts to detect trends in population size of chamois *Rupicapra rupicapra* and *R. pyrenaica*?, *Wildlife Biology*, 12, 1, 77-88.
- Lorimer J., 2008. Counting corncrakes. The affective science of the UK corncrake census, *Social Studies of Science*, 38, 3, 377-405.
- Maris V., 2018. *La part sauvage du monde*, Paris, Seuil.
- Meine C., Soulé M., Noss R., 2006. A mission-driven discipline: the growth of conservation biology, *Conservation Biology*, 20, 3, 631-651.
- Riopel A., 2018. Des phoques océanographes en Antarctique, *Le Monde*, 6 juin, supplément Science et Médecine.
- Sandbrook C., 2015. The social implications of using drones for biodiversity conservation, *Ambio*, 44, Suppl. 4, S636-S647.
- Sarkar I.N., 2009. Biodiversity informatics: the emergence of a field, *BMC Bioinformatics*, 10, Suppl. 14, S1.
- Soberon J., Peterson A.T., 2004. Biodiversity informatics: managing and applying primary biodiversity data, *Philosophical Transactions of the Royal Society B*, 359, 1444, 689-698.
- Strasser B.J., 2011. Data-driven sciences: from wonder cabinets to electronic databases, *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, 1, 85-87.
- Strasser B., Edwards P.N., 2017. Big data is the answer. But what is the question?, *Osiris*, 32, 1, 328-345.
- Sutherland W.J., Pullin A.S., Dolman P.M., Knight T.M., 2004. The need for evidence-based conservation, *Trends in Ecology and Evolution*, 19, 6, 305-308.
- Turnhout E., Boonman-Berson S., 2011. Databases, scaling practices, and the globalization of biodiversity, *Ecology and Society*, 16, 1, 35.
- Zimmerman A.S., 2008. New knowledge from old data. The role of standards in the sharing and reuse of ecological data, *Science, Technology and Human Values*, 33, 5, 631-652.

Citation de l'article : Arpin I. Les producteurs professionnels de données sur la biodiversité face aux « big data » en écologie. *Nat. Sci. Soc.* 28, 1, 66-72.