

LA SCIENCE EN RÉSEAU

Les gestionnaires d'information « invisibles » dans la production d'une base de données scientifiques

[Florence Millerand](#)

S.A.C. | « *Revue d'anthropologie des connaissances* »

2012/1 Vol. 6, n° 1 | pages 163 à 190

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-anthropologie-des-connaissances-2012-1-page-163.htm>

Distribution électronique Cairn.info pour S.A.C..

© S.A.C.. Tous droits réservés pour tous pays.

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

LA SCIENCE EN RÉSEAU

Les gestionnaires d'information « invisibles » dans la production d'une base de données scientifiques

FLORENCE MILLERAND

RÉSUMÉ

Les développements technologiques contemporains en matière d'infrastructures numériques permettent d'observer de nouvelles formes de travail, voire de nouvelles catégories de travailleurs, dans les milieux scientifiques. Basé sur une étude ethnographique d'un réseau américain de chercheurs en écologie, cet article s'attache au travail des gestionnaires d'information (*information managers*), « techniciens invisibles » responsables de la gestion des données scientifiques au sein des laboratoires. Il montre comment le développement d'un projet de très grande base de données s'accompagne de processus de mise en visibilité et invisibilité des gestionnaires d'information et de leur travail, et discute des enjeux de la mise en invisibilité des activités de documentation des données en particulier, sur les processus de production des connaissances scientifiques. L'invisibilité des gestionnaires d'information apparaît reliée à un aspect fondamental de leur travail, en l'occurrence un « travail d'articulation » caractérisé par des activités de bricolage, de traduction et d'effacement. Le travail des petites mains de la gestion et de la documentation des données dans les sciences est sans cesse à refaire et à réinventer.

Mots clés : travail invisible, technicien invisible, gestionnaire d'information, metadonnée, standard, infrastructure

INTRODUCTION

Les milieux scientifiques connaissent actuellement des développements technologiques majeurs en matière d'infrastructures numériques basées sur les réseaux, qui visent, entre autres, la circulation et le partage des données de recherche entre les frontières institutionnelles et disciplinaires. À l'instar des premières grandes banques de données en sciences de la vie (cf. la GenBank ou la Protein Data Bank), un mouvement similaire se développe dans d'autres domaines scientifiques, notamment en sciences de la nature où les défis scientifiques actuels (le changement climatique ou la biodiversité par exemple) requièrent la mise en commun de grandes quantités de données pluridisciplinaires. On voit ainsi émerger la constitution de gigantesques entrepôts de données, juxtaposant de façon inédite des champs disciplinaires auparavant jamais rapprochés et donnant à observer de nouvelles formes de travail, voire de nouveaux travailleurs, dans les milieux scientifiques.

La question de la fabrication de ces nouveaux équipements scientifiques est encore peu explorée. Elle apparaît pourtant cruciale. La façon dont nous classons, enregistrons et stockons les données, informations et connaissances (que ce soit dans des notes manuscrites, des livres ou des bases de données) façonne inextricablement la façon dont nous connaissons (Goody, 1979 ; Bowker, 2006). Bowker a montré comment la construction de bases de données dans le domaine de la biodiversité revenait à organiser et à classer le monde naturel d'une façon qui excluait inéluctablement certains espaces, certaines entités et temporalités (Bowker, 2000). Nous avons montré comment le choix d'un standard informatique pour le stockage et le partage de données en sciences de l'environnement revenait à privilégier *a priori* une perspective disciplinaire sur une autre, en facilitant l'accès à des données environnementales physiques plutôt qu'à des données biologiques (Millerand et Bowker, 2008, 2009). Au-delà de leurs enjeux épistémiques et scientifiques, les systèmes de classifications, les normes et conventions terminologiques peuvent également rendre visibles (ou invisibles) des catégories d'acteurs et d'activités, et ce faisant, produire un certain ordonnancement du monde – par exemple, via l'instauration d'une certaine division du travail, associée à des mécanismes de reconnaissance sociale et professionnelle (Bowker et Star, 1999 ; Desrosières et Thévenot, 1988 ; Thévenot, 1986). Il s'agit dès lors de se pencher sur le travail concret de fabrication de ces infrastructures, un travail d'arrière-scène, en coulisse, pour saisir le rôle de la matérialité des supports d'une part, et la diversité des acteurs et des activités en jeu d'autre part.

Cet article s'attache à une catégorie d'activités et d'acteurs peu habitués à être sous les projecteurs, que l'on regroupe dans la catégorie fourre-tout du travail technique et des « techniciens ». Il s'agit plus précisément du travail des gestionnaires d'information (*information manager*), c'est-à-dire des personnes qui, au sein des laboratoires et groupes de recherche, s'occupent de la gestion des données de recherche, ce qui inclut généralement des activités de traitement, de sauvegarde, de stockage, voire de collecte sur le terrain et

d'analyse. À partir d'une étude empirique basée sur un réseau américain de chercheurs en écologie (ci-après « le Réseau » avec un grand R), cet article vise à mieux comprendre la participation des gestionnaires d'information à l'activité scientifique, et plus précisément à envisager les processus de mise en visibilité ou invisibilité du travail de ces « techniciens invisibles » (Shapin, 1989), dans le contexte particulier du développement d'une grande base de données. Nous montrerons comment, dans le cas étudié, la constitution d'une base de données à l'échelle du Réseau s'est accompagnée de processus de mise en visibilité et invisibilité des gestionnaires d'information et de leur travail, et comment cela a contribué à redéfinir partiellement leur rôle au sein du Réseau.

Nous commençons par présenter les éléments théoriques et méthodologiques qui permettent de situer notre recherche. Nous exposons ensuite l'étude empirique et le terrain de recherche sur lesquels sont basés les arguments développés dans cet article. Nous développons alors un argumentaire en deux volets. Dans un premier temps, nous décrivons le travail des gestionnaires d'information, en insistant plus particulièrement sur l'activité de documentation des données scientifiques. Nous proposons une explication de l'invisibilité des gestionnaires d'information et de leur travail en la reliant à un aspect fondamental de leur travail que nous appréhendons en tant qu'un « travail d'articulation » au sens de Strauss (1985, 1988, 1992) et que nous caractérisons, dans le contexte particulier de la gestion des données scientifiques, par des activités de bricolage, de traduction et d'effacement. Dans un deuxième temps, nous montrons comment, dans le cours du développement de la base de données, les gestionnaires de données deviennent à la fois partiellement « visibles » – dans la mesure où on leur reconnaît un nouveau rôle et de nouvelles responsabilités – et « invisibles » – dans la mesure où une grande partie de leur travail fait l'objet d'une mise en invisibilité. Nous discutons des enjeux de la mise en invisibilité des activités de documentation des données en particulier, sur les processus de production des connaissances scientifiques. Nous concluons en suggérant des pistes d'analyse pour la prise en compte et l'étude du travail des techniciens dans les développements contemporains d'infrastructures numériques dans les sciences.

TECHNICIENS, INFRASTRUCTURES ET TRAVAIL INVISIBLE

Dans les laboratoires, les techniciens sont juste des techniciens. Dans l'histoire officielle des réalisations scientifiques, ils n'existent ni par leur nom, race, genre ou identité mais seulement par leur fonction (Timmermans, 2003, p. 197).

Les techniciens de laboratoire, petites mains qui travaillent à faire avancer la science aux côtés des chercheurs, ont longtemps été absents, non seulement

des récits historiques des découvertes scientifiques (Shapin, 1989), mais aussi des travaux des sociologues des sciences. Récemment, plusieurs travaux ont ouvert la voie à des questionnements de recherche sur le travail de ces « petites mains », acteurs souvent peu visibles dans la production et la circulation des connaissances, qu'il s'agisse de la contribution des amateurs (ou des « non-scientifiques ») (Mukerji, 2009 ; Epstein, 1995) ou des techniciens au sens large (Barley et Bechky, 1994 ; Shapin, 1989 ; Cambrosio et Keating, 1988 ; Lynch, 1985 ; Collins, 1974 ; Goodwin, 1995 ; Timmermans, 2003). Les études de laboratoire ont attiré l'attention sur le travail pratique de fabrication des faits scientifiques. Ce faisant, elles ont souligné le rôle d'acteurs intermédiaires, dont les techniciens, qui travaillent à faire le lien entre les entités du monde physique (la mesure d'une quantité de carbone) et le monde symbolique (la biomasse d'une forêt), entre les particularités d'une expérimentation locale et un système de connaissance (Latour et Woolgar, 1979 ; Traweek, 1988 ; Lynch, 1985 ; Knorr-Cetina, 1981). Cependant, si on leur reconnaît un rôle clef dans les opérations de « traduction », notamment dans la production des « inscriptions » (Latour et Woolgar, 1979), la contribution des techniciens est rarement analysée en tant que telle (du point de vue d'une sociologie du travail scientifique), mais presque toujours en fonction de leur contribution aux processus de production des connaissances scientifiques (du point de vue d'une sociologie de la connaissance) (Barley et Bechky, 1994). Au final, les travaux qui se sont penchés sur le rôle des techniciens dans les sciences à partir d'un questionnement sur l'organisation sociale du travail dans les laboratoires restent peu nombreux.

Pour les non-techniciens, les connaissances et savoir-faire des techniciens sont généralement perçus comme étant complexes, voire hermétiques, et leur rôle comme étant avant tout un second rôle, au service d'autres tâches, plus intellectuelles, qui occupent le premier rang de l'activité scientifique, comme la rédaction d'articles scientifiques (Barley et Bechky, 1994). Par ailleurs, les milieux scientifiques auraient une propension à dévaluer les collaborateurs au statut social inférieur (Shapin, 1989), non seulement parce que ceux-ci assument une fonction de « soutien » (les techniciens travaillant au service des chercheurs qu'ils assistent), mais aussi parce que les connaissances qu'ils détiennent sont essentiellement des connaissances dites contextuelles, certes indispensables à la production scientifique, mais dont le prestige est moindre par rapport aux connaissances dites formelles (Barley et Bechky, 1994). De ce fait, les techniciens – comme la nature précise du travail technique – restent, le plus souvent, « invisibles » (Star et Strauss, 1999).

Et pourtant, les techniciens assument en pratique une diversité de tâches dans le cadre desquelles ils mobilisent un vaste ensemble de savoirs, que ce soit sur le plan théorique ou pratique. Ainsi, Lynch (1982, 1985) montre comment les techniciens de laboratoire en neurobiologie qui manipulent les microscopes électroniques, mobilisent à la fois des connaissances théoriques (donc certaines théories scientifiques) et des savoirs situés pour identifier les anomalies et artefacts – que les neurobiologistes sont incapables de repérer eux-mêmes.

Dans un autre registre, Orr (1996) montre comment les techniciens spécialisés dans la réparation de photocopieurs présentent des niveaux de connaissances pointues et sophistiquées sur les copieurs, qui sont similaires à celles des ingénieurs qui les ont conçus.

Les techniciens sont, au sein des laboratoires, les acteurs les plus proches des instruments et des données, c'est-à-dire de la recherche sur le terrain (ou sur la paillasse) où le travail quotidien implique un contact étroit avec les infrastructures matérielles du travail scientifique. Au-delà de la sphère scientifique, la notion d'infrastructure matérielle renvoie communément à l'idée de vastes systèmes technologiques (comme le réseau de chemin de fer ou le système d'électricité), qui sont généralement transparents pour l'usager, considérés à l'arrière-plan des activités sociales et qui ne deviennent visibles qu'en cas de pannes (Star et Ruhleder, 1996). Il est intéressant de dresser un parallèle avec le travail des techniciens chargés de la maintenance de tels systèmes qui, lui aussi, n'apparaît généralement qu'en cas de problèmes. Leur performance sera considérée d'autant meilleure que le résultat de leur travail restera invisible. Les infrastructures numériques contemporaines (dont Internet constitue la figure emblématique) n'échappent pas à la règle.

L'étude de telles infrastructures dans le contexte des milieux scientifiques pousse à élargir l'unité d'analyse – jusque-là centrée sur le fait scientifique ou le dispositif technique – à l'infrastructure sociotechnique appréhendée en tant que « système de production de connaissances » (Star et Ruhleder, 1996). Il s'agit de centrer le questionnement sur les processus de construction des infrastructures d'une part, et prendre en compte la matérialité des choses d'autre part. Star (1999) nous suggère d'aller voir en coulisse (« *going backstage* ») et de pratiquer ce que Bowker appelle « l'inversion infrastructurelle » (Bowker, 1994), c'est-à-dire rapporter au premier plan ces infrastructures traditionnellement prises pour acquies. Dans cette perspective, l'étude du travail accompli en pratique ou « en action » (Latour, 1987) plutôt que l'étude des discours d'acteurs rend possible la mise à jour des écarts existants entre ceux qui réalisent le travail et ceux qui en sont récompensés.

Par ailleurs, l'attention à l'infrastructure matérielle, aux objets et à leur « équipement » (Vinck, 2006), peut permettre d'accéder à des acteurs ou activités qui n'apparaissent pas toujours clairement dans les discours officiels ou spontanés (Vinck, 2009). Les propriétés des objets, la façon dont on les « équipe » (Vinck, 2009), matérialisent des infrastructures invisibles faites de standards, de catégories, de conventions (Bowker et Star, 1999) qui, tout en permettant leur circulation, la contraignent inévitablement. Choisir certaines catégories et certains formats pour décrire des données scientifiques revient inéluctablement à en exclure d'autres, par exemple lorsqu'on indique le nom du laboratoire plutôt que celui du chercheur comme auteur des données. Les bases de données sont aussi des lieux de représentations et de luttes politiques (Bowker, 2000 ; Meyer, 2009 ; Waterton, 2010).

Nous prenons appui sur les notions de « travail visible » et « travail invisible », formulées par Strauss dans ses études sur la maladie et les milieux hospitaliers

(Strauss, 1988, 1992 ; Strauss et al., 1985). Ces notions ont été reprises, entre autres, par Star et Bowker (1999), dans leur travail sur les enjeux épistémiques et politiques des systèmes de classification (Bowker et Star, 1999), et par Suchman (1995) sur les systèmes de représentation en informatique. Le « travail invisible » renvoie, le plus souvent, à un ensemble d'activités mal comprises (les tâches informelles de diffusion d'information dans une organisation, par exemple) ou à une catégorie d'acteurs peu considérée, généralement en bas de l'échelle dans les hiérarchies sociales et professionnelles (le travail des domestiques de maison par exemple) (Nardi et Engeström, 1999). Mais il peut aussi s'agir d'activités de travail hautement spécialisées et abstraites, qui impliquent des tâches de gestion d'information ou de connaissances (le travail de gestion des données de recherche dans un laboratoire, par exemple).

Nous adoptons la posture selon laquelle « aucun travail n'est en soit visible ou invisible ; nous « voyons » le travail à travers une sélection d'indicateurs » (Star et Strauss, 1999, p. 9). Une activité peut *devenir* visible ou invisible en fonction du contexte, notamment des rapports de pouvoir en place. La tension entre le visible et l'invisible est l'objet de négociations constantes (Star et Strauss, 1999). La mise en visibilité d'une activité de travail peut mener à la reconnaissance sociale du travail et du travailleur, mais elle peut aussi conduire à une réification du travail et ouvrir la voie à de nouvelles formes de contrôle (Suchman, 1995). Dans le cas des infirmières étudiées par Bowker et al. (1995), les tentatives de mise en visibilité de leur travail (et plus largement de leur profession) se sont heurtées à la nécessité de maintenir le caractère ambigu et personnel de certaines tâches, par exemple l'accompagnement d'un mourant dans ses dernières minutes de vie. Enfin, si la mise en visibilité du travail implique nécessairement son objectivation et une certaine appréciation de sa valeur *a priori*, il reste qu'une grande partie échappe à la codification formelle. Toute mise en visibilité du travail maintient une part d'invisible irréductible (Voirol, 2005).

Une piste pour penser l'articulation entre le visible et l'invisible consiste à se questionner sur la façon dont on définit le travail, sur ce qui compte comme étant du travail et, pour qui, en fonction de qui, ce travail peut – ou doit – être visible ou au contraire invisible. La notion de « travail d'articulation » développée par Strauss (1985, 1988, 1992) s'oppose à celle de routine, c'est-à-dire à ce qui fonctionne parce que préalablement ordonné, fixé, délimité et qui peut être pris pour acquis. Le travail d'articulation fait référence à l'ensemble des activités qui font tenir ensemble une série de tâches, d'acteurs et d'environnements. Pour reprendre les termes de Strauss, c'est « ce qui fait en sorte que le travail peut être accompli » (Strauss, 1988). Concrètement, ce travail d'articulation repose essentiellement sur des activités (largement invisibles) de planification, d'organisation, d'évaluation, de médiation et de coordination de tâches, qui permettent d'« articuler » un ensemble de composantes (tâches, individus, technologies, environnements, etc.). Rendre le travail d'articulation visible, c'est reconnaître la primauté du processus, de l'ambigu et de l'émergent sur ce qui est produit, quantifiable et prévisible. C'est aussi reconnaître la valeur de

ceux et celles qui en ont la charge et dont le rôle est souvent mal compris et sous estimé (Star, 1991a). À ce titre, les fonctions de coordination en général en constituent un exemple éclairant.

Choisir de rendre compte du travail d'une catégorie de travailleurs sous représentée dans les travaux sociologiques sur l'activité scientifique contribue inéluctablement à la visibilité d'un point de vue en particulier (Star, 1991b). Ce faisant, il devient possible d'expliquer différemment les cas de succès ou d'échec des innovations techniques et de mettre à jour des processus de mise en visibilité/invisibilité des personnes et des activités. Le travail de laboratoire est aussi une activité morale et politique où toutes les contributions d'acteurs ne sont pas évaluées de la même façon, selon les statuts (Shapin, 1989), la race (Timmermans, 2003) ou le genre (Garforth et Kerr, 2010).

LE CAS DES « GESTIONNAIRES D'INFORMATION » DANS UN RÉSEAU SCIENTIFIQUE

Les arguments développés dans cet article reposent sur une étude ethnographique réalisée au sein d'un réseau américain de chercheurs en écologie. Né au début des années 1980, le Réseau représente la plus grande communauté scientifique dans le domaine avec plus de 2000 chercheurs et étudiants répartis dans 26 sites (ou stations de recherche), auxquels s'ajoute un 27^e site chargé de l'administration du Réseau. Fondamentalement multidisciplinaire, le Réseau croise plus de 60 champs disciplinaires, de la climatologie à la zoologie en passant par la biologie végétale, la géologie ou la limnologie. Chaque site regroupe une équipe d'une douzaine de chercheurs (en moyenne) autour de l'étude d'un biome particulier (un ensemble d'écosystèmes propres à une région donnée, par exemple un désert chaud ou un estuaire côtier). On retrouve ainsi, au sein de chacun des sites, des cultures épistémiques qui peuvent varier considérablement. En effet, même si les écologues du Réseau partagent une certaine vision de la Nature et de l'étude des écosystèmes, leur expertise diffère largement en fonction de leurs objets et méthodes de recherche (phytoplancton, pingouins, flux de nutriments, etc).

L'archivage des données de recherche est une priorité du Réseau depuis sa création. Il faut dire que le domaine des sciences de l'environnement se caractérise par la manipulation de très grandes quantités de données (de terrain notamment). Ainsi, un poste est dédié aux tâches de gestion des données de recherche au sein de chacun des sites¹. C'est un « gestionnaire d'information » (aussi appelé « gestionnaire de données »), le plus souvent formé à la fois en informatique et en sciences de l'environnement, qui occupe la fonction

1 Ce poste est prévu dans le financement du Réseau depuis sa création.

(seul ou en équipe dans les plus gros sites). Environ un tiers des gestionnaires d'information sont aussi des docteurs es sciences (en informatique ou en sciences de la nature), les autres sont diplômés de premier ou de deuxième cycle universitaire. La plupart des gestionnaires d'information qui ont un doctorat sont actifs dans des projets de recherche, au sein du Réseau ou dans d'autres communautés de recherche.

Largement distribués sur le plan géographique², les gestionnaires d'information forment en quelque sorte une communauté de pratique. Ils se réunissent chaque année dans le cadre du congrès annuel du Réseau où on leur réserve une journée de présentations, ateliers, démonstrations, etc. Ils sont amenés à communiquer et à collaborer fréquemment dans le cadre de groupes de travail (par télé ou visioconférence), et ils disposent de leur propre bulletin d'information. Bien que rassemblant des profils très diversifiés, la communauté des gestionnaires d'information du Réseau se distingue par son identité forte et par sa stabilité dans le temps. Sur le plan organisationnel, ils sont représentés par un comité propre (*Information Management Committee*) et au sein du conseil d'administration du Réseau.

Le gestionnaire d'information est en charge de toutes les données de recherche produites au sein de son site de recherche. Il veille à la coordination de l'ensemble des tâches reliées à la production, la manipulation, l'archivage et, dans certains cas, à l'exploitation des données : collecte sur le terrain, traitement, vérification de la qualité, stockage, analyse, publication des données sur le site Web et dans les bases de données, etc. Concrètement, une fois recueillies sur le terrain, les données « brutes » sont transmises au gestionnaire d'information qui les numérise (si ce n'est pas déjà fait), les vérifie, et les enregistre dans une base de données. Ce faisant, il doit, entre autres, *documenter* les données (ou compléter les documentations si celles-ci ont déjà été faites), c'est-à-dire inscrire des métadonnées (littéralement des « données sur les données ») de façon à ce que le contexte de production des données soit conservé.

Documenter les données consiste à indiquer des informations générales telles que le titre du projet de recherche ou le nom du ou des chercheurs, et des informations plus détaillées sur les techniques d'échantillonnage, les instruments, les unités de mesure, etc. Généralement, une partie de ces informations est consignée sur le site même de recueil de données, sur papier ou tablette électronique, par les chercheurs ou les personnels. Ces informations seront ensuite largement élaborées (complétées et vérifiées) par les gestionnaires d'information, au moment de l'enregistrement et du traitement des données sur support informatique (base de données) dans le laboratoire. La documentation des données revêt une importance toute particulière dans le contexte du Réseau, car celui-ci s'est engagé, depuis les années 2000, à rendre public l'ensemble des

2 Les 26 sites et stations de recherche du Réseau sont répartis partout aux États-Unis, en Antarctique et en Arctique.

données produites au sein de ses 26 sites de recherche³. En d'autres termes, les données de chacun des chercheurs et de chacun des sites du Réseau doivent être disponibles, en ligne, à des fins de partage et de réutilisation, non seulement au sein du Réseau mais aussi à l'extérieur. Ceci implique, en premier lieu, que les documentations de données soient suffisamment complètes et détaillées pour pouvoir franchir les murs des laboratoires – et des traditions disciplinaires – dans lesquels elles ont été produites. Ces documentations correspondent à ce que Vinck (2011) appelle l'équipement, c'est-à-dire l'addition d'une chose à un objet qui en modifie le statut et les propriétés et qui, notamment, l'inscrit dans un espace d'échange et de circulation. À titre d'exemple, une série de mesures chimiques sur l'eau d'un milieu aquatique donné ne pourra être comparée à une autre que si les détails des conditions de prélèvement (date, localisation géographique, type d'instrument, niveau de profondeur, technique d'analyse...) les accompagnent. Les documentations (métadonnées) constituent la condition *sine qua non* de la mobilité des données scientifiques. C'est essentiellement à cette tâche de documentation, aussi appelée contextualisation des données ou production de métadonnées que nous allons nous intéresser dans le cadre de cet article.

La documentation des données constitue, à l'heure actuelle, une activité importante – sinon la plus importante – du travail des gestionnaires d'information du Réseau. Il faut dire que la politique du Réseau en matière de libre circulation des données de recherche prend appui sur un projet de développement d'une très grande base de données, lui-même basé sur un processus de standardisation des procédés de documentation de données. En effet, le projet de libre circulation des données dans le Réseau s'est rapidement heurté à la grande diversité des systèmes, architectures techniques, modes de classement et de documentation des données qu'on y trouve. Un standard commun de description de données (ou standard de métadonnées⁴) a alors été adopté (en 2001) pour uniformiser les façons de faire entre les 26 sites (la figure 1 présente un exemple de métadonnées standardisées). Le projet de « cyberinfrastructure » (terme utilisé par le Réseau pour le projet de très grande base de données) vise, non seulement la publication des données à des fins de partage entre chercheurs, mais aussi, la conduite, en ligne, d'analyses fines sur de très grands ensembles de données pluridisciplinaires, ce qui requiert un niveau de documentation de données extrêmement fin.

3 Les sites disposent d'un délai maximum de trois ans pour la mise en ligne de l'ensemble de leurs données à partir du moment où celles-ci ont été recueillies. Cette exigence est intégrée à l'exercice d'évaluation de la production scientifique des sites qui a lieu tous les trois ans et qui détermine le maintien du financement des activités de recherche.

4 Le standard, appelé « *Ecological Metadata Language* », est un langage standardisé de description de données développé spécialement pour le domaine de l'écologie.

```

<geographicCoverage>
  <geographicDescription>Ficity, Fl, metropolitan area,
    USA</geographicDescription>
  <boundingCoordinates>
    <westBoundingCoordinate>-112.373614</westBoundingCoordinate>
    <eastBoundingCoordinate>-111.612936</eastBoundingCoordinate>
    <northBoundingCoordinate>33.708829</northBoundingCoordinate>
    <southBoundingCoordinate>33.298975</southBoundingCoordinate>
    <boundingAltitudes>
      <altitudeMinimum>300</altitudeMinimum>
      <altitudeMaximum>600</altitudeMaximum>
      <altitudeUnits>meter</altitudeUnits>
    </boundingAltitudes>
  </boundingCoordinates>
</geographicCoverage>
<temporalCoverage>
  <rangeOfDates>
    <beginDate>
      <calendarDate>1998-11-12</calendarDate>
    </beginDate>
    <endDate>
      <calendarDate>2003-12-31</calendarDate>
    </endDate>
  </rangeOfDates>
</temporalCoverage>

```

Figure 1. Description de coordonnées géographiques et temporelles dans le langage du standard de métadonnée

En pratique, la mise en œuvre du standard commun de description de données, qui constitue la pierre d'assise du projet de très grande base de données, s'est avérée un processus extrêmement complexe et laborieux, marqué par la confrontation des « cultures de données » propres aux différentes équipes de recherche⁵. En 2007, soit six ans après l'adoption du standard commun, à peine la moitié des sites avait réussi à documenter leurs données de façon standardisée, et il aura fallu attendre 2009 pour que, finalement, l'ensemble des sites ait généré leurs premiers jeux de données proprement documentés. Au moment d'écrire ces lignes, le projet de grande base de données est toujours en développement.

Dans le cadre de notre étude ethnographique, nous avons suivi ce projet de base de données en prêtant une attention particulière au processus de standardisation des méthodes de documentation de données qui l'a accompagné. En suivant le principe de « l'inversion infrastructurelle » de Bowker (1994), nous avons observé le travail concret de développement, à la fois d'une infrastructure technique (décisions et choix en matière d'architecture, de standards techniques, etc.) et de son contenu (types de données, champs disciplinaires représentés, etc.), un travail au sein duquel les gestionnaires d'information se sont avérés être, *en pratique*, des acteurs de premier plan, et dans le cadre duquel leur travail a fait l'objet d'une certaine redéfinition, notamment à travers des processus de mises en visibilité et d'invisibilité. L'enquête repose sur un travail d'observation participante, d'entretiens avec les acteurs, d'analyses de documents (documentation technique, rapports,

5 Sur ce point, voir Baker et Millerand (2010).

etc.) et d'artefacts (bases de données, prototypes, etc.) sur une période de deux ans (de 2004 à 2006). Nous étions basés à l'institution universitaire où deux des 26 sites de recherche sont établis. Nous avons pu observer, *in situ* et au quotidien, le travail des chercheurs, étudiants, professionnels de recherche, techniciens, en prêtant une attention particulière aux gestionnaires d'information. Nous avons observé ces sites en participant à leurs activités et aux événements qui s'y déroulaient : situations de travail, conférences scientifiques, réunions, conférences téléphoniques, ateliers de conception (*design sessions*), conversations de couloir, etc.

PROCESSUS DE MISE EN VISIBILITÉ/ INVISIBILITÉ

La documentation des données scientifiques : un travail de petites mains

Susan est gestionnaire d'information dans un des 26 sites du Réseau, au sein duquel elle travaille avec deux techniciens programmeurs à la gestion des données de recherche. L'équipe de recherche du site rassemble une dizaine de chercheurs, principalement des océanologues, autour de l'étude des effets des changements climatiques sur les écosystèmes marins ; une vingtaine d'étudiants et postdoctorants ainsi qu'une dizaine de personnels (essentiellement des techniciens de terrain (« *field technician* ») complètent l'équipe. Nous décrivons ici le travail de documentation de données à partir d'un exemple de recueil de données dans le cadre d'une mission scientifique en mer.

La mesure de la biomasse (masse des organismes vivants dans un milieu donné) est à la base des recherches menées par l'équipe de recherche du site pour mesurer les changements dans l'écosystème. Les prélèvements se font en mer, dans le cadre de missions ponctuelles, sur plusieurs jours ou semaines, à bord d'un navire scientifique. Avant chaque mission, Susan prépare les formulaires de documentation de données sur lesquels, une fois en mer, les chercheurs consigneront les informations relatives aux prélèvements et mesures effectués à bord. Les formulaires papier, fixés sur une tablette, sont remplis par les chercheurs, étudiants et techniciens qui se relaient à tour de rôle à la fois sur le pont pour manipuler les instruments et collecter les échantillons, et dans le mini-lab en cabine pour réaliser les premières analyses. Susan reste, quant à elle, à terre au laboratoire. Elle nous explique les détails du processus de collecte de données qui servent à mesurer la biomasse⁶ :

6 La mesure d'un taux de concentration de chlorophylle est utilisée ici comme un indicateur de la quantité de plancton végétal dans un milieu aquatique, qui correspond à la biomasse (ou masse des organismes vivants dans un milieu).

D'abord, on a le fluorimètre c'est l'instrument utilisé pour mesurer la biomasse ou plutôt la chlorophylle en mer. Donc l'instrument on l'a calibré et on l'amène en mer. En mer, les bouteilles d'échantillonnage (*sampling bottle*) sont descendues sur le côté, de façon à ce que l'eau soit captée à différentes profondeurs, l'eau est alors remontée dans de grandes bouteilles, de 15 litres, et puis on utilise des petites bouteilles de 200 millilitres pour prélever des sous échantillons. Après on prend une bouteille de chaque et on l'apporte au lab, là on a un dispositif de filtrage dans lequel on insère un filtre d'une certaine taille et on verse notre échantillon dans le filtre. Tout le plancton qui contient la chlorophylle reste pris dans le filtre. Ensuite on prend ce filtre, on le met au freezer 24 h de façon à ce que les cellules se séparent, ensuite on prend ces échantillons gelés et on les met dans de l'acétone pour extraire la chlorophylle des cellules (...) on obtient un liquide qu'on va mettre dans le fluorimètre (...) ça va nous permettre de dire que la fluorescence va être proportionnelle à la chlorophylle.

Résumé de cette façon, le protocole de collecte apparaît relativement simple : les prélèvements d'eau de mer sont effectués en utilisant des bouteilles plongées à différentes profondeurs, le contenu est ensuite filtré et analysé grâce à un appareil appelé fluorimètre. Il s'agit d'une manipulation relativement connue en planctologie. Cependant, dès lors qu'il s'agit de documenter les données recueillies, ici des concentrations de chlorophylle, une série de questions se pose : dans quel contexte les données sont-elles recueillies (pour quel projet de recherche, à quelle date, par quelle équipe, etc.) ?, quelle est la technique d'échantillonnage utilisée (des bouteilles de quelle contenance, à quels niveaux de profondeur, etc.) ?, avec quelle technique d'extraction (filtrage, de quel type, taille, etc.) ?, quelle méthode d'analyse (fluorescence, pigments, etc.) ? L'ensemble des réponses à ces questions peut représenter un grand nombre d'informations, généralement supérieur, en quantité, aux données elles-mêmes.

Le processus de documentation des données ne débute pas *ex nihilo* à partir du moment où l'équipe quitte le terrain (ou le navire), mais il commence en mer, au moment même où les instruments sont déployés et lors des premières analyses, dans le mini-lab en cabine :

D'abord quand tu recueilles tes échantillons, tu commences à utiliser la feuille, et tu écris la profondeur, la station, latitude, longitude, etc. Donc les échantillons ont désormais des labels qui indiquent son contexte de recueil. Après, quand tu fais tourner le fluorimètre, pour chaque lecture, tu écris sur la ligne appropriée sur la feuille, la fluorescence, disons la lecture se fait en unités de fluorescence. (...) On écrit aussi les caractéristiques de l'instrument de mesure (...). Tu dois écrire aussi combien d'eau a été filtrée, etc.

Les premières informations consignées dans le formulaire servent à documenter le contexte de collecte des données dans le temps et dans l'espace (date, localisation géographique, profondeur, nom du chercheur, etc.). Il s'agit des métadonnées de premier niveau. Ensuite seront indiqués les

volumes des échantillons analysés, les valeurs mesurées, les caractéristiques de l'instrument de mesure, etc. ; ces informations constitueront les métadonnées de deuxième niveau. Les métadonnées de premier niveau servent essentiellement à identifier des ensembles de données, par exemple à partir d'une recherche sur le sujet (ex. : biomasse), tandis que les métadonnées de deuxième niveau permettent leur interprétation et analyse. Sans les détails de la méthode d'analyse par exemple, les données auraient peu d'intérêt pour un chercheur qui souhaiterait les exploiter.

Les notes de ce bloc-notes vont ensuite être entrées sur Excel, dans ce cas ils font le calcul en mer, ou alors ils vont attendre et entrer les données en rentrant au lab, des fois c'est juste un fichier texte, pas Excel, ça dépend. (...) Quand les données reviennent à terre, l'équipe phytoplancton va travailler un peu sur les données et ensuite, ils vont nous envoyer le fichier, généralement par courriel.

Une fois remplis, les formulaires papier sont transférés sur support informatique et passent des mains de l'équipe de recherche à celles de la gestionnaire d'information. Ces formulaires jouent un rôle particulièrement important dans la coordination du travail au sein de l'équipe de recherche. Créés par la gestionnaire d'information, ils sont apportés avec le matériel sur le navire où ils circulent entre le pont et le mini-lab, passant de main en main, pour ensuite retourner à la gestionnaire d'information. En circulant entre les différents lieux de l'activité de recherche (en mer sur le navire et sur terre dans le laboratoire) et entre les différents acteurs (chercheurs, étudiants, techniciens, gestionnaire d'information), ils s'apparentent aux objets-frontière décrits par Star et Griesemer (1989) qui servent de support à la coordination entre différents mondes sociaux. Plus précisément, le formulaire agit ici comme une « forme standardisée », servant à la fois de méthode de communication entre les différents acteurs et de mémoire (ou dépôt d'informations) au service de la conservation des données.

L'attention aux modalités de circulation des formulaires révèle le caractère distribué du processus de documentation des données, qui est réalisé en pratique par plusieurs catégories d'acteurs. Par ailleurs, les documents utilisés sont des documents intermédiaires, voués à se transformer et à disparaître (le formulaire papier devient un fichier électronique, lui-même sera intégré dans un formulaire modèle). La distribution de l'activité et le caractère éphémère des supports utilisés rendent difficile la conservation de traces d'une part, et l'identification de rôles bien définis d'autre part, deux facteurs qui contribuent à l'invisibilité de ces tâches dans l'ensemble du processus de recherche.

Une fois qu'on a le fichier, on doit le mettre dans un formulaire « modèle », un gabarit. Parce que des fois [en mer] ils changent l'ordre des colonnes parce que c'est plus facile pour eux de les remplir, des fois ils ajoutent des colonnes, donc on doit les vérifier. Par exemple, ils vont avoir enregistré les latitudes en degrés, minutes, secondes, c'est généralement comme ça qu'on fait parce que c'est ce qui est le plus simple à comprendre. Donc

nous, on doit convertir ces trois colonnes en degrés décimaux, parce que c'est comme ça que les données peuvent être interopérables. Aussi, par exemple, s'ils nous ont envoyé les données chlorophylle avec trois points de décimales, nous on sait que c'est en fait un point décimal, les trois points c'est à cause de l'instrument parce que c'est digital, donc on vérifie quand même avec eux, mais on corrige. (...) Après, une fois qu'on a complété les métadonnées, on fait passer les données de notre environnement de production [base de données du site à accès privé] à la sphère publique, on change les privilèges et donc là tout est public et consultable.

Le fichier Excel en main, le travail de la gestionnaire d'information consiste alors à en réviser le contenu, à le compléter et, surtout, à standardiser les documentations produites, par exemple en utilisant des modèles de formulaires prédéfinis. Le travail évoqué ici en est un de traduction (on convertit les formats personnels dans un format standard) mais aussi de correction (on remet les colonnes en ordre, on ajuste les valeurs, on convient d'un nombre maximum de décimale, etc.). En d'autres mots, on travaille à effacer les traces des contextes particuliers dans lesquels les données ont été recueillies (ce que Latour et Woolgar (1979) ont appelé, dans un autre contexte, le processus d'effacement des modalités) pour laisser place à des données épurées, dénuées des idiosyncrasies propres aux personnes et aux cultures de laboratoire. Les gestionnaires d'information réalisent ainsi un véritable travail de polissage, visant à rendre les données mobiles, indépendantes des environnements dans lesquels elles ont été produites et pourtant pleines des précieuses informations documentaires qui permettront de les re-situer dans leurs contextes d'origine.

Dans ce processus, le gestionnaire d'information joue un rôle clef car il intervient directement sur la qualité des données produites, en permettant leur interopérabilité (et donc leur comparaison avec d'autres), en corrigeant les erreurs, anomalies ou artefacts introduits par les instruments et en ajoutant aux données brutes les précieuses informations contextuelles qui permettront leur interprétation. Son niveau d'expertise dans le domaine de recherche entre en jeu dans la qualité des données et des documentations produites. Or, comme l'explique Susan, « le problème, c'est que nous les gestionnaires d'information, on est des traducteurs donc on peut prendre ça littéralement (...) les chercheurs eux ils savent donc ils corrigent dans leur tête (enfin on espère) ». Il s'agit là d'un enjeu important au sein du Réseau, car le gestionnaire d'information ne peut être formé à l'ensemble des méthodes utilisées par les chercheurs de son site.

L'attention au travail particulier de documentation des données scientifiques permet de mieux saisir la nature du travail des gestionnaires d'information qui s'apparente, plus largement, à un travail d'« articulation » (Strauss, 1988).

« On ne fait pas des choses qui sont dans les critères d'évaluation que la communauté valorise. On n'écrit pas des demandes de subvention à dix millions. On ne publie pas des milliards de papiers tous les ans. On est

trop occupés juste à *faire en sorte que le travail se fasse*. Alors sur la base de ces critères que la plupart des communautés scientifiques valorisent traditionnellement, on est tout simplement *invisible*⁷. » [c'est nous qui soulignons]

« Faire en sorte que le travail se fasse » décrit précisément ce qui est au cœur du travail d'articulation qui caractérise certaines activités de travail dans les organisations. Il s'agit d'un travail d'arrière-plan, généralement mal compris, non formalisé et difficilement formalisable, un travail, à proprement parlé, *invisible*, qui fait que le travail se fait et que les individus et les collectifs travaillent et coopèrent ensemble.

Dans son étude sur l'organisation du travail médical à l'hôpital (Strauss et al, 1985), Strauss montre comment les infirmières ont en charge l'articulation des différentes perspectives d'acteurs (médecins, patients, familles, etc.), un travail qui consiste en des tâches de supervision et de coordination fondamentales à l'administration des soins, mais qui reste méconnu, ignoré et peu valorisé. De la même façon, les gestionnaires d'information sont au cœur de relations de collaboration qui engagent différentes catégories d'acteurs (chercheurs, personnels, étudiants...) et différentes communautés scientifiques, au sein desquelles ils réalisent un travail invisible de coordination et de médiation. Or, comme l'illustre l'extrait cité ci-dessus, les gestionnaires d'information sont invisibles aux yeux des chercheurs aussi et surtout parce que leur travail ne correspond pas à celui que la communauté scientifique valorise et qu'il ne peut être ni mesuré ni quantifié à l'aune des critères d'évaluation traditionnels de la communauté scientifique (nombre de publications ou montants des subventions).

Cela étant dit, le travail des gestionnaires d'information se caractérise également par une série d'activités de *bricolage* (marquées par des savoirs situés), de *traduction* (entre les domaines technique et scientifique, et entre des mondes sociaux) et d'*effacement* (des traces de l'activité), qui contribuent tout autant à leur invisibilité.

Lévi-Strauss opposait le bricoleur, sorte d'homme à tout faire dont les compétences et les connaissances sont généralement mises en œuvre en fonction du contexte, à l'ingénieur dont les outils et les aptitudes sont mis au service de la matérialisation de projets bien particuliers. À bien des égards, les gestionnaires d'information sont des bricoleurs. Ils apprennent les rudiments du métier sur le tas, ils doivent faire preuve d'une grande polyvalence et l'improvisation est leur lot quotidien. Le travail de documentation de données lui-même peut s'apparenter à du bricolage. Ainsi, un des gestionnaires d'information du Réseau raconte comment, plusieurs années avant l'adoption du standard, il avait créé de toutes pièces un format de documentation de données au nom évocateur de « *Everybody Happy Format, EHF* » (que l'on pourrait traduire par « le format sur

7 Citation tirée de l'étude ethnographique réalisée en 2004 par Karasti et Baker au sein du Réseau, avec la permission des auteurs. Voir Karasti et Baker (2004).

lequel tout le monde s'entend ») pour satisfaire les besoins de son site. En tenant compte des contraintes techniques informatiques (souci d'interopérabilité et de maintenance), des contraintes scientifiques (compatibilité avec les autres documentations en vigueur dans les domaines disciplinaires connexes) et des attentes de chacun des chercheurs de son site (sur les types de données à décrire), il avait « bricolé » un format « maison » qui, à l'époque, satisfaisait les principaux acteurs concernés.

Les gestionnaires d'information sont aussi des traducteurs, en ce sens qu'ils réalisent des opérations de traduction entre les aspects techniques et scientifiques et entre des mondes sociaux. La première dimension de traduction est sans doute la plus tangible et correspond à la fonction même du technicien, spécialiste d'un certain domaine technique. Les gestionnaires d'information convertissent des données de recherche « brutes » dans des formats qui les rendent lisibles et interprétables par les chercheurs. La deuxième dimension renvoie à des opérations de traduction entre des mondes sociaux. Les gestionnaires d'information sont au cœur d'un réseau de relations qui croisent de multiples domaines disciplinaires, organisations et institutions. Si les chercheurs rattachés au site auquel appartient le gestionnaire constituent ses principaux interlocuteurs, celui-ci est amené à collaborer avec l'ensemble des chercheurs du Réseau, avec d'autres communautés de recherche et avec le public (dans la mesure où les données sont accessibles sur le Web). Ce faisant, les gestionnaires d'information traversent constamment des frontières disciplinaires (par exemple, lorsqu'il s'agit d'intégrer dans une même base des données biologiques, chimiques ou physiques) ou institutionnelles (par exemple, lorsqu'il s'agit de publier des données du Réseau dans une archive institutionnelle d'une université en particulier).

Enfin, la nature même du travail des gestionnaires d'information consiste à faire en sorte d'en effacer toutes les traces et de le rendre « transparent ». Les gestionnaires d'information travaillent très fort pour masquer le fouillis de données désordonnées, mal présentées, pleines d'artefacts avec lesquelles ils travaillent, et pour confiner le travail assommant de « nettoyage », d'organisation et de classement des données à l'intérieur du système. C'est là un aspect important du travail des techniciens, comme le montre Orr (1996) dans son ethnographie des techniciens réparateurs de photocopieurs qui prennent soin d'ôter toute tâche de poudre qui pourrait témoigner de leur passage, ou encore Suchman (1995) dans son étude sur les techniciens documentalistes dans les cabinets d'avocats dont les résultats du travail doivent se fondre dans les documents présentés en cour. Le travail des gestionnaires d'information dans le processus de documentation de données peut rarement être lu ou regardé : les formulaires de documentation papier que les chercheurs emportent avec eux sur le terrain constituent les rares documents visibles du processus. Une fois déplacés sur support électronique, ils sont intégrés à un système de classification et encastés dans une architecture de base de données qui restera, pour l'essentiel, invisible aux yeux de ses utilisateurs.

La mise en visibilité des gestionnaires d'information : la reconnaissance d'un nouveau rôle

Le travail de standardisation, de vérification et de complétion des documentations de données scientifiques est au cœur du travail des gestionnaires d'information qui nous préoccupe dans cet article. C'est précisément ce processus de standardisation que le Réseau a décidé de pousser à un niveau supérieur, en adoptant un standard commun de métadonnées dans le cadre de son projet de très grande base de données. Il s'agit, non plus seulement que les documentations de données soient standardisées à l'échelle des équipes de recherche (au sein d'un même site), mais qu'elles le soient aussi à l'échelle du Réseau tout entier, voire de la communauté de recherche en écologie.

L'adoption du standard vise deux changements majeurs dans les pratiques de documentation de données. D'une part, toute documentation doit désormais respecter le langage et la structure du standard, ce qui implique que les sites doivent adopter de nouvelles façons de faire. D'autre part, les chercheurs sont censés être les principaux acteurs de ce processus, en assumant désormais eux-mêmes la tâche de documentation des données.

En pratique, la production de métadonnées standardisées s'avèrera être un processus extrêmement laborieux et, face à la complexité de la tâche, les chercheurs rejeteront quasiment immédiatement ce nouveau rôle, qui reviendra aux gestionnaires d'information. Ce faisant, la complexité du travail de documentation, et ceux qui en seront ultimement responsables, c'est-à-dire les gestionnaires d'information, bénéficieront d'une certaine forme de « visibilité ».

Le sale boulot

Lorsque le Réseau adopte le standard en 2001, les usagers présumés ou « configurés » (pour reprendre l'expression de Woolgar, 1991) sont les chercheurs. À ce stade, aucune autre catégorie d'acteurs (techniciens ou gestionnaires d'information par exemple) n'est mentionnée dans les documents relatifs au standard, destiné à « favoriser le partage des données entre les *chercheurs* » qui sauront mettre à profit ce tout nouveau standard représentant « le dernier cri (*state-of-the-art*) des standards de métadonnées »⁸. On s'attend à ce que les chercheurs, qui connaissent leurs données mieux que n'importe qui d'autre, les documentent et qu'ils les publient directement dans la grande base de données.

Or les chercheurs du Réseau vont d'emblée refuser la tâche, d'une part en raison de la quantité de travail impliquée (et de sa complexité), d'autre part parce qu'ils estiment qu'elle relève du travail de gestion des données au sens large, un travail habituellement assumé par les gestionnaires d'information.

« La quantité de métadonnées à produire est dix fois plus importante que la taille des données collectées ! Et puis, les chercheurs, ils vont pas s'y mettre... Ils préfèrent faire leur recherche ».

8 Document interne ; c'est nous qui soulignons.

Si, comme l'indique cette gestionnaire d'information, la lourdeur de la tâche rebute d'emblée les chercheurs, il s'agit aussi et surtout de tâches que les chercheurs cherchent généralement à éviter, voire qu'ils ne veulent même pas voir.

« Tu sais, au fond, c'est parce qu'ils ne veulent pas avoir à s'en occuper, alors ils sont très contents que je puisse les débarrasser de ça... ».

Les chercheurs ayant refusé le rôle qu'il leur était imparti dans le « script » (Akrich, 1992) du projet de base de données, la tâche revient rapidement et presque « naturellement » aux gestionnaires d'information (mot d'un interviewé). Cette délégation de la tâche aux gestionnaires d'information s'apparente à ce que Hughes (1996) a appelé la délégation du sale boulot (« *dirty work* ») dans les soins de santé, en l'occurrence des médecins vers les infirmières. Il constatait que, lorsqu'une profession était amenée à déléguer des tâches, c'était souvent des tâches que celle-ci jugeait en bas de l'échelle des valeurs sociales, fastidieuses voire dégradantes, et que la délégation visait généralement à réaffirmer une hiérarchisation des compétences, afin de préserver un certain prestige. Dans le cas qui nous occupe, les chercheurs du Réseau ont rapidement trouvé en la personne du gestionnaire d'information de leur site, un moyen de se « débarrasser » du problème, d'autant plus que la justification était toute trouvée : celui-ci s'occupait déjà du travail de gestion des données.

Les gestionnaires d'information commencent alors le travail, de conversion des documentations existantes dans un premier temps, et de production des documentations standardisées dans un deuxième temps. Cependant, à l'image de la pluridisciplinarité qui caractérise le Réseau, les pratiques de documentation sont extrêmement disparates et « résistent » aux efforts de standardisation entrepris. Premièrement, le standard s'avère complexe ; s'approprier cette nouvelle façon de décrire les données revient littéralement à s'approprier un nouveau langage (en l'occurrence les bases du langage XML) avec lequel il faut « coder » les données. Or les gestionnaires d'information n'ont ni l'expertise ni la capacité de trouver les ressources additionnelles nécessaires (ne serait-ce qu'en termes de temps disponible). Ils se heurtent par ailleurs à un manque flagrant de soutien au sein du Réseau, comme l'évoque ce gestionnaire :

« Une de mes plus grandes frustrations, c'est qu'on nous demande de le faire sans nous accorder aucun financement supplémentaire alors que ça exige des ressources énormes. »

Dans ce contexte, rares sont ceux qui parviennent à recruter du personnel supplémentaire (un stagiaire par exemple). La majorité d'entre eux n'ont d'autres choix que de se former sur le tas. Dans tous les cas, l'expérience est pénible : ils se sentent dépourvus, « incompetents » ou « sous-qualifiés » (dans les mots des interviewés). Deuxièmement, tant les outils techniques destinés à faciliter la mise en œuvre du standard (notamment les outils de conversion automatisée des documentations existantes vers des documentations standards)

que l'encadrement fourni (en termes de formations) s'avèrent inutiles ou inappropriés. Les outils posent d'importants problèmes de compatibilité et les formations offertes sont en décalage avec les besoins des sites (les courbes d'apprentissage variant en fonction des formations de base des gestionnaires d'information). Troisièmement, et c'est là que les difficultés se font plus pointues, le standard pose des problèmes de fond qui dépassent les responsabilités des gestionnaires d'information en touchant aux aspects épistémiques des données de recherche. Très rapidement, les gestionnaires réalisent que le fait de changer la façon dont ils documentent les données entraîne des conséquences importantes sur la façon dont les données sont enregistrées, archivées, voire même produites. À titre d'exemple, lorsque les systèmes de classification sont propres à un laboratoire voire à un chercheur, faut-il les corriger ? Si non, comment les « faire rentrer » dans les catégories prédéfinies du standard ?

Façonner les catégories

« Nous, on utilise beaucoup d'unités personnalisées dans nos projets, [des unités] qu'on a créées nous-mêmes. (...) Ces unités-là, elles n'ont pas d'équivalence dans le standard, alors on est obligé de les mettre en vrac dans le champ "méthode" ... au fond ça ne sert pas à grand-chose. »

Dans l'extrait cité, la gestionnaire d'information évoque une absence, en l'occurrence une catégorie manquante pour les « unités personnalisées », qu'elle est « obligée de mettre en vrac dans le champ "méthode" ». La raison pour laquelle « ça ne sert pas à grand-chose », c'est que la catégorie « méthode » est une sorte de tiroir « fourre-tout » dans lequel les méthodes sont décrites sous la forme d'un texte. Le niveau de précision et de complétude de ces descriptions peut varier considérablement puisqu'il est laissé à la discrétion de l'auteur de s'assurer que l'ensemble des informations importantes y est consigné. Dans ces circonstances, la qualité des informations est extrêmement variable.

Dans leur travail sur les systèmes de classification, Bowker et Star (1999) montrent la force invisible et contraignante des catégories dans la production d'un ordre social, qui peuvent contribuer à privilégier certains points de vue ou au contraire à les réduire au silence. Ce problème du manque d'équivalence pour les unités personnalisées évoqué dans l'extrait ci-dessus renvoie à la tension entre les deux traditions disciplinaires dominantes au sein du Réseau, en l'occurrence entre les disciplines qui relèvent de la physique (ex. : la climatologie) et celles qui relèvent de la biologie (ex. : la biologie marine), deux traditions qui renvoient à des types de données très différents⁹. Les données physiques utilisent des unités de mesure standard (ex. : degrés Celcius pour la température, mètres pour la profondeur, etc.), alors que les données

9 Notons que, dans le champ scientifique, l'écologie fait partie des sciences biologiques. Cependant, elle fait aussi appel à plusieurs disciplines des sciences physiques, notamment la géographie physique, la géologie, la pédologie, etc.

biologiques utilisent, en plus des unités standards, des unités « personnalisées », créées pour les besoins particuliers du projet (ex. : « nombre de feuilles sur le tronçon inférieur », « proportion d'œufs vivants versus morts », etc.). Or il se trouve que le standard adopté par le Réseau est plutôt inspiré de modèles en sciences physiques, et donc qu'il est plus aisé d'y « faire rentrer » des données physiques¹⁰. Il faut alors, dans le cas de données biologiques, ou bien « tordre » les catégories existantes, ou bien en façonner de nouvelles.

Ce travail de façonnage des catégories va occuper les gestionnaires d'information dans l'ensemble des sites et va entraîner plusieurs changements dans le standard. Susan et son équipe aboutiront par exemple à la création d'un ensemble de nouvelles sous-catégories destinées à structurer le champ « méthode », qui permettra notamment de capturer la diversité des techniques d'échantillonnage et des protocoles d'analyse. En permettant de préciser les méthodologies, ces nouvelles catégories permettent aussi et surtout de compléter les documentations de façon à ce que les données puissent être interprétées correctement, ce qui renvoie directement à la question de la qualité des données de recherche.

Par exemple, le traitement de l'échantillon, c'est par extraction ou pigmentation ? Si c'est par extraction, quelle taille de filtre ?... Si on n'a pas le type de filtre ou sa taille, alors c'est un vrai problème, parce que, si tu prends des jeux de données et que tu les regroupes alors que, dans un cas, c'est tel filtre et dans l'autre, tel autre, tu introduis des erreurs majeures dans l'analyse, et c'est ce qui arrive...

Dans l'extrait cité, l'omission de la taille du filtre utilisé dans la technique d'extraction peut conduire à amalgamer des données de façon erronée et mener ainsi à la production de résultats faussés.

Bon, je sais que Margaret utilise souvent le filtre HA et Rob le GFF, tous les deux ont l'habitude d'utiliser des filtres particuliers mais ils ne pensent jamais à l'indiquer, et ils ne précisent pas la taille parce qu'ils oublient et souvent ils utilisent qu'une seule taille donc ils oublient qu'il y en a plusieurs (...) quand je sais pas je vais les voir et là ils tombent des nues, ça les embête parce que eux ils se posent pas la question.

Le travail de documentation introduit de nouvelles contraintes, qui confrontent directement les façons de faire des chercheurs et des équipes de recherche, voire les cultures des laboratoires. Margaret ne précise pas les détails du dispositif qu'elle utilise (nom et taille de filtre) parce qu'il s'agit d'un savoir implicite (elle n'utilise que ce filtre en particulier). La tâche des gestionnaires d'information s'apparente alors à un véritable processus d'élicitation de connaissances visant à expliciter les techniques d'échantillonnage, les méthodes d'analyse, les unités de mesure, etc. Les savoirs sont situés et intimement liés

¹⁰ Cet aspect, et plus largement les enjeux politiques des bases de données, ont été discutés dans Millerand et Bowker (2008, 2009).

aux cultures épistémiques. Le façonnage des catégories du standard force l'explicitation des connaissances, discute le choix des conventions, tend à formaliser les savoirs tacites. S'amorce alors un nouveau type d'échange entre les gestionnaires d'information et les chercheurs, dans le cadre duquel se discutent des choix de dénomination, de classification, d'organisation et de présentation des données. L'exercice permettra de rendre visibles, aux yeux des chercheurs en particulier, les difficultés auxquelles les gestionnaires d'information sont confrontés, ainsi que le travail concret d'élaboration du contenu du standard.

Un nouveau rôle

En 2005, lors du congrès annuel des gestionnaires d'information et alors que l'effort de standardisation arrive dans sa quatrième année, les gestionnaires d'information se voient reconnaître un nouveau rôle, celui de « co-développeurs » du standard. Pour la première fois, leur nom apparaît aux côtés de celui du standard. Les gestionnaires d'information sont reconnus non seulement à l'échelle du Réseau, mais plus largement à l'échelle de la communauté scientifique en écologie. Ils deviennent *visibles* :

« [Le standard] est défini et révisé dans le cadre d'un effort continu de la communauté scientifique, auquel participent en particulier les gestionnaires d'information du Réseau¹¹. »

Cette reconnaissance intervient dans la foulée d'une série de sondages sur la progression du processus de standardisation dans les sites, et suite à une rencontre marquante entre les gestionnaires d'information du Réseau et les développeurs du standard, qui prendra la forme d'un atelier dans le cadre du congrès annuel de 2005. Ces deux événements vont constituer des mécanismes de mise en visibilité du travail réalisé par les gestionnaires d'information. D'un côté, les sondages révéleront l'ampleur et la complexité de la tâche, en révélant notamment les retards dans le processus de standardisation. Ce faisant, ils contribuent à construire le processus de standardisation en tant que phénomène (Igo, 2007), ici un processus complexe et laborieux. De l'autre, l'atelier sera l'occasion, pour les gestionnaires d'information, de montrer les défauts et les limites inhérentes au standard et, surtout, les « corrections » qu'ils y ont apportées. À cette occasion, ils offriront des démos, présenteront leurs outils, discuteront des différentes solutions, etc. L'atelier donnera lieu à la publication d'un compte rendu dans le bulletin trimestriel des gestionnaires d'information du Réseau, co-signé par les représentants des deux parties (gestionnaires et auteurs du standard). Le rôle des gestionnaires d'information en tant que co-développeurs du standard est désormais reconnu sur le plan institutionnel.

11 Document interne.

La mise en invisibilité des gestionnaires d'information : le déni de reconnaissance des processus de travail

La reconnaissance institutionnelle des gestionnaires d'information se fera cependant au détriment de la reconnaissance de certains processus qui, pourtant fondamentaux dans le travail de documentation de données, seront gardés invisibles.

La visibilité ou l'invisibilité d'une activité de travail repose généralement sur des procédés d'inscription et de formalisation des tâches, ce que Voirol (2005) appelle la « visibilité formelle » dans les milieux de travail. Ainsi, on utilisera des dispositifs de formalisation pour rendre « visible » une activité de travail, par exemple une description de tâches associée à un titre, un niveau de revenu, de responsabilités, etc. L'écrit joue, dans ce registre, un rôle prépondérant. Les gestionnaires d'information développeront une série d'outils et de compléments au standard, dont des « bonnes pratiques » (*Best Practices*) destinées à faciliter la production de métadonnées standardisées. Le document, qui se veut être un guide pratique visant à faciliter le travail de documentation de données, est, aussi et surtout, une tentative de formalisation, par écrit, de nouveaux processus de travail. Cependant, il sera rejeté par les auteurs du standard et ignoré par les administrateurs du Réseau. L'activité pratique de production de documentations de données standardisées restera ainsi sans trace écrite reconnue.

La rédaction du document commence dès 2003, au moment où une poignée de gestionnaires d'information suggère de rédiger un document commun qui expliquerait, de façon simple, le contenu du standard. Après plusieurs mois de travail, le document, qui a parcouru la plupart des sites du Réseau, rassemble, entre autres, une série de définitions et d'explications sur ce que doit inclure ou ne pas inclure tel ou tel champ. À titre d'exemple, chaque jeu de données doit être lié à un « projet », au même titre qu'il est associé au nom d'un chercheur ou d'un laboratoire. Malgré son univocité apparente, la catégorie « projet » s'est avérée tout à fait polysémique en regard des différents contextes disciplinaires des sites.

« Je me souviens en train de me demander : est-ce que le “projet” c'est la “mission” (*cruise*) donc pour nous dans notre système “l'étude” (*study*) ou est-ce que c'est le nom du “site” ? On peut interpréter ça de plusieurs manières, par exemple dans le site de Rob, admettons qu'ils travaillent sur les champignons en ce moment, donc pour eux, “projet”, ce sera le nom du projet : “Projet champignons 2002-2004”. Mais si nous on faisait ça avec nos missions, comme on en a plusieurs dans un même projet, on aurait en fait plusieurs projets : “Étude de janvier 2001”, “Étude de mars 2001”... (*Jan01-Study, March01-Study*) alors qu'elles sont réalisées dans le cadre d'un même grand projet de recherche. Donc ça crée des problèmes parce que les données ne peuvent pas être comparées à partir du moment où elles ne sont pas organisées de la même façon. »

L'exemple cité ici renvoie à des modes de classement et d'organisation des données qui sont intimement liés à des traditions de recherche. Univoque et non problématique en biologie végétale, où il est facile d'associer un recueil de données à un projet de recherche, le recueil de données en océanographie rend leur classement autrement plus complexe. Les croisières scientifiques (les missions) sont en effet l'occasion de collecter des données sur plusieurs projets (pour des raisons évidentes de coût) mêlant ainsi des mesures chimiques, biologiques ou physiques. Les données produites forment alors une collection en tant que telle (ex. : « Étude de janvier 2001 »), où chaque mission est définie comme une entité : « chaque mission est une histoire en tant que telle » (dans les mots d'une interviewée). On comprend alors que de telles différences puissent constituer des obstacles importants à la mise en commun de données, notamment à travers une base de données. C'est précisément de ces différences, et de leurs conséquences sur la production de documentations standardisées, dont rend compte le document sur les bonnes pratiques. Pour revenir à l'exemple cité, il sera ainsi précisé que, dans le contexte de recherches pluridisciplinaires incluant notamment des jeux de données océanographiques, ceux-ci devraient être associés d'abord à l'équipe ou au site de recherche qui les a produits, et ensuite à un ou plusieurs projets de recherche.

Les « bonnes pratiques » rédigées par les gestionnaires d'information sont plus qu'une série de recommandations techniques, elles sont aussi l'expression de processus de travail formalisés, qui attribuent des rôles à des catégories d'acteurs (ex. : le gestionnaire d'information est responsable de vérifier les métadonnées avec l'auteur des données), des ordonnancements de tâches (ex. : suivre la feuille de route sur les étapes de conversion), qui traduisent des responsabilités ou des manquements à certaines responsabilités, par exemple ici une documentation déficiente du standard en ce qui concerne les données pluridisciplinaires. Plus encore, le document rend compte de tout le travail d'interprétation et de façonnage des catégories réalisé par les gestionnaires d'information.

Voyant dans cet effort de formalisation une preuve tangible de leur contribution à l'amélioration du standard, et avec l'idée que d'autres communautés pourraient bénéficier de leur expérience, les gestionnaires d'information proposent d'inclure ces « bonnes pratiques » à la documentation livrée avec le standard. Mais celles-ci seront disqualifiées d'emblée par les auteurs du standard qui en jugeront le contenu trop « provincial » (dans les mots d'un interviewé). Considérées trop « amateurs » et trop empruntées des particularités locales des sites, les « bonnes pratiques » rédigées par les gestionnaires d'information s'inscrivent en porte-à-faux avec la vocation universelle du standard, comme l'évoque un développeur :

« Les bonnes pratiques auraient pu faire partie de la documentation du standard si elles avaient été plus générales, parce qu'on veut desservir la communauté en écologie la plus large possible. »

En rejetant l'idée d'intégrer le document à la documentation du standard, le travail des gestionnaires d'information est rendu *invisible*. Plus précisément, ce qui est rendu invisible, c'est le travail local, situé et diversifié, de fabrication de documentations standardisées, c'est-à-dire un travail d'interprétation et de façonnage de catégories pour y « faire rentrer » des cultures de laboratoire et des traditions disciplinaires. Non formalisés, ces savoirs sur la fabrication des documentations, le façonnage des catégories, etc., pourront difficilement faire l'objet de processus d'accumulation, et les autres communautés n'auront d'autres choix que d'emprunter le même chemin que les gestionnaires d'information du Réseau. Mais il faut dire aussi qu'accepter de reconnaître ce travail de petites mains reviendrait à écorcher le caractère universel du standard et, par là, à remettre en cause la logique de développement de haut en bas (*top-down*) du Centre national de recherche qui a créé un standard destiné à solutionner les problèmes de documentation de données de toutes les communautés de recherche en écologie. Au final, le travail des petites mains de la documentation des données est constamment à refaire et à réinventer.

CONCLUSION

Les développements actuels en matière d'infrastructures de collaboration scientifique, notamment de grandes bases de données, qui mettent au jour de nouvelles formalisations des pratiques de recherche, invitent à une meilleure compréhension du rôle des acteurs qui travaillent dans les coulisses de la production scientifique. L'analyse des mécanismes de mise en visibilité ou invisibilité, des processus de travail et des travailleurs, dans le cadre d'un projet de grande base de données au sein d'une communauté scientifique montre comment certaines facettes du travail d'une catégorie d'acteurs, en l'occurrence les gestionnaires d'information, sont révélées ou au contraire ignorées. On assiste ainsi à un processus d'« invisibilisation » (Denis, 2009) qui n'est pas seulement le résultat d'une disparition des traces de l'activité de travail, mais aussi une forme de dévalorisation de l'activité en tant que telle.

Cela étant dit, si, comme les techniciens, « leur travail constitue la couche la plus facile à effacer du palimpseste de la science » (Timmermans, 2003, p. 198), les gestionnaires d'information pourraient voir leur rôle prendre une nouvelle importance au moment où les projets de grandes bases de données scientifiques se multiplient. À cheval entre la science informatique et les sciences de l'environnement, ces travailleurs invisibles détiennent des savoir-faire et des connaissances qui en font un maillon indispensable à la production et à la circulation de la connaissance scientifique, notamment via les réseaux numériques. À l'ère des initiatives de type « science en réseau », passer outre le travail des « petites mains » des gestionnaires d'information conduirait nécessairement à une compréhension appauvrie des pratiques scientifiques contemporaines.

Remerciements

Je tiens à remercier Karen Baker pour sa précieuse collaboration, ainsi que les éditeurs du numéro et les évaluateurs anonymes qui ont grandement contribué à l'amélioration de cet article. Je remercie également les membres du Réseau pour avoir accepté de participer à cette recherche, ainsi que le Conseil de recherche en sciences humaines du Canada (CRSH) pour l'avoir financée.

RÉFÉRENCES

- Akrich M. (1992). The de-description of technical objects. In W. E. Bijker et J. Law (Eds.), *Shaping technology/Building Society. Studies in Sociotechnical Change*. Cambridge, MA: MIT Press, pp. 259-263.
- Baker, K.S., & Millerand, F. (2010). Infrastructuring ecology: New challenges for data sharing. In J. Parker, N. Vermeulen & B. Penders (Eds.), *Collaboration in the new life sciences*. Surrey. Burlington: Ashgate, pp. 111-138.
- Barley S. R. & Bechky B. A. (1994). In the Backrooms of Science: The Work of Technicians in Science Labs. *Work and Occupations: An International Sociological Journal*, 21 (1), 85-126
- Bowker G. C. (2006). *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.
- Bowker G. C. (2000). Biodiversity Datadiversity. *Social Studies of Science*, 30 (5), 643-684.
- Bowker G. C. (1994). *Science on the Run: Information Management and Industrial Geophysics at Schlumberger, 1920-1940*. Cambridge, MA: MIT Press.
- Bowker G. C. & Star S. L. (1999). *Sorting things out: classification and its consequences*. Cambridge, Mass.: MIT Press.
- Bowker G. C. Timmerman S., & Star S. L. (1995). Infrastructure and organizational transformation: classifying nurses' work. In W. Orlikowski, G. Walsham, M. R. Jones, & J. I. DeGross (Eds.), *Information technology and changes in organizational work*. London: Chapman & Hall, pp. 344-370.
- Cambrosio A. & Keating P. (1988). "Going Monoclonal": Art, Science, and Magic in the Day-to-Day Use of Hybridoma Technology. *Social Problems, Special Issue: The Sociology of Science and Technology*, 35 (3), 244-260.
- Collins H. M. (1974). The TEA Set: Tacit Knowledge and Scientific Networks. *Science Studies*, 4, 165-186.
- Denis J. (2009). Le travail invisible de l'information. In C. Licoppe (Ed.), *L'évolution des cultures numériques, de la mutation du lien social à l'organisation du travail*. Paris: FYP, pp. 117-123.
- Desrosières A. & Thévenot, L. (1988). *Les catégories socioprofessionnelles*. Paris: La Découverte.
- Epstein S. (1995). The construction of lay expertise: AIDS activism and the forging of credibility in the reform of clinical trials. *Science, Technology & Human Values*, 20 (4), 408-413.
- Forsythe D. E. (1993). Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence. *Social Studies of Science*, 23 (3), 445-477.
- Garforth L. & Kerr A. (2010). Let's Get Organised: Practicing and Valuing Scientific Work Inside and Outside the Laboratory. *Sociological Research Online*, 15 (2), <<http://www.socresonline.org.uk/15/12/11.html>>
- Goodwin C. (1995). Seeing in Depth. *Social Studies of Science*, 25, 237-274.
- Goody J. (1979). *La raison graphique. La domestication de la pensée sauvage*. Paris : Éditions de Minuit.

- Honneth A. (2005). Invisibilité : sur l'épistémologie de la « reconnaissance ». *Réseaux*, 1 (129/130), 39-57.
- Hughes E.-C. (1996). *Le regard sociologique. Essais choisis*. Paris : Éditions de l'EHESS.
- Igo S. E. (2007). *The Averaged American: Surveys, Citizens, and the Making of a Mass Public*. Cambridge-London: Harvard University Press.
- Karasti H. & Baker K. S. (2004). *The Long-Term Information Management Trajectory: Working to Support Data, Science and Technology*. Scripps Institution of Oceanography Technical Report: University of California San Diego.
- Knorr-Cetina, K. (1981). *The manufacture of knowledge. An essay on the constructivist and contextual nature of science*. Oxford: Pergamon Press.
- Knorr-Cetina K. (1999). *Epistemic Cultures. How the Sciences Make Knowledge*. Cambridge: Harvard University Press.
- Knorr-Cetina K. (2001). Laboratory Studies: Historical Perspectives. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences*. Amsterdam-New York: Elsevier, pp. 8232-8238.
- Latour B. (1987). *Science in action: how to follow scientists and engineers through society*. Cambridge, Mass.: Harvard University Press.
- Latour B. & Woolgar S. (1979). *Laboratory Life: The Social Construction of Scientific Facts*. Los Angeles, CA: Sage.
- Law J. (1994). *Organizing Modernity*. Oxford: Blackwell.
- Lynch M. (1985). *Art and artifact in laboratory science: a study of shop work and shop talk in a research laboratory*. London: Routledge & Kegan Paul.
- Lynch M. (1982). Technical Work and Critical Inquiry: Investigations in a Scientific Laboratory. *Social Studies of Science*, 12 (4), 499-534.
- Meyer M. (2009). Objet-frontière ou Projet-frontière ? Construction, (non-)utilisation et politique d'une banque de données. *Revue d'anthropologie des connaissances*, 3(1), 128-148.
- Millerand F. & Bowker, G. C. (2009). Metadata Standards. Trajectories and Enactment in the Life of an Ontology. In S. L. Star & M. Lampland (Eds.), *Standards and Their Stories*. Cornell University Press, pp. 149-165.
- Millerand F. & Bowker G. C. (2008). Metadata, trajectoires et « éaction ». In C. Rosental (Ed.), *La cognition au prisme des sciences sociales*. Paris : Éditions des Archives contemporaines, pp. 277-303.
- Mukerji C. (2009). *Impossible engineering: technology and territoriality on the Canal du Midi*. Princeton, NJ: Princeton University Press.
- Nardi B. A. & Engeström Y. (1999). A Web on the Wind: The Structure of Invisible Work. *Computer Supported Cooperative Work*, 8 (1/2), 1-8.
- Orr J. E. (1996). *Talking about Machines: An Ethnography of a Modern Job*. New York, NY: Cornell University Press.
- Roth W. M. & Bowen G. M. (2001). Of disciplined minds and disciplined bodies: On becoming an ecologist. *Qualitative Sociology*, 24 (4), 459-481.
- Shapin S. (1989). The Invisible Technician. *American Scientist*, 77, 554-563.
- Star S. L. (1999). The Ethnography of Infrastructure. *American Behavioral Scientist* (43), 377-391.
- Star S. L. (Ed.) (1995). *Ecologies of Knowledge. Work and Politics in Science and Technology*. Albany, NY: State University of New York.
- Star S. L. (1991a). The sociology of the invisible: The primacy of work in the writings of Anselm Strauss. In D. R. Maines (Ed.), *Social Organization and Social Process*. New York: Aldine de Gruyter, pp. 265-283.
- Star S. L. (1991b). Power, Technologies and the Phenomenology of Standards: On Being Allergic to Onions. In J. Law (Ed.), *A Sociology of Monsters? Power, Technology and the Modern World*. London: Routledge, pp. 27-57.
- Star S. L. (1989). The Structure of Ill-Structured Solutions: Heterogeneous Problem-Solving, Boundary Objects and Distributed Artificial Intelligence. In M. Huhns, L. Gasser (Eds.), *Distributed Artificial Intelligence*. Menlo Park: Morgan Kaufmann, pp. 37-54.
- Star S. L. & Griesemer J. (1989). Institutional Ecology, 'Translations,' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-1939. *Social Studies of Science*, 19, 387-420.

- Star S. L. & Ruhleder K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Informations Systems Research*, 7 (1), 111-134. Version française dans *Revue d'Anthropologie des Connaissances*, 4(1), 114-161, 2010.
- Star S. L. & Strauss A. (1999). Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *JCSCW*, 8 (1/2), 9-30.
- Strauss A. (1985). Work and the Division of Labor. *The Sociological Quarterly*, 26, 1-19.
- Strauss A. (1988). The Articulation of Project Work: An Organizational Process. *The Sociological Quarterly*, 29 (2), 163-178.
- Strauss A. (1992). *La trame de la négociation. Sociologie qualitative et interactionnisme*. Paris : L'Harmattan.
- Strauss A., Fagerhaugh S., Suczek B., & Wiener, C. (1985). *Social Organization of Medical Work*. Chicago, IL: University of Chicago Press.
- Suchman L. A. (1995). Making Work Visible. *Communications of the ACM: Special Issue on Representations of Work*, 38 (9), 56-64.
- Thévenot L. (1986). Les investissements de forme. In L. Thévenot (Ed.), *Conventions économiques*. Paris : Presses universitaires de France (Cahiers de Centre d'Étude de l'Emploi), pp. 21-71.
- Timmermans S. (2003). A Black Technician and Blue Babies. *Social Studies of Science*, 33 (2), 197-229.
- Traweek S. (1988). *Beamtimes and lifetimes: the world of high energy physicists*. Cambridge, MA: Harvard University Press.
- Vinck D. (2011). Taking intermediary objects and equipping work into account in the study of engineering practices, *Engineering Studies*, 3 (1), 25-44.
- Vinck D. (2009). De l'objet intermédiaire à l'objet-frontière. Vers la prise en compte du travail d'équipement. *Revue d'anthropologie des connaissances*, 3 (1), 51-72.
- Vinck D. (2006). L'équipement du chercheur : comme si la technique était déterminante. *ethnographiques.org [en ligne]*, 9, <http://www.ethnographiques.org/documents/article/ArVinck.html>.
- Voirol O. (2005). Présentation. Visibilité et invisibilité : une introduction. *Réseaux*, 1 (129/130), 9-36.
- Waterton C. (2010). Experimenting with the Archive: STS-ers As Analysts and Co-constructors of Databases and Other Archival Forms. *Science, Technology & Human Values*, 35 (5), 645-676.
- Woolgar S. (1991). Configuring the user: the case of usability trials. In S. Law (Ed.), *A Sociology of Monsters: Essays on Power, Technology and Domination*. London: Routledge, pp. 57-99.

Florence MILLERAND est professeure au Département de communication sociale et publique à l'Université du Québec à Montréal (UQAM). Elle co-dirige le Laboratoire de communication médiatisée par ordinateur (LabCMO) et elle est membre du Centre interuniversitaire de recherche sur la science et la technologie (CIRST). Ses recherches se situent au croisement des études en communication et des études en science, technologie, société (STS). Elles portent sur les technologies de communication, le Web social et les infrastructures de collaboration scientifique (e-science).

Affiliation	Département de communication sociale et publique Université du Québec à Montréal Case postale 8888, succursale Centre-ville Montréal (Québec) H3C 3P8 CANADA
Courriel	millerand.florence@uqam.ca

ABSTRACT : NETWORK SCIENCE. THE « INVISIBLE » INFORMATION MANAGERS IN THE PRODUCTION OF A SCIENTIFIC DATABASE

Contemporary network-based technological developments in the sciences draw attention to new forms of work, or even new categories of workers. Based on an ethnographic study of an American ecological research network, this paper focuses on the work of information managers, the « invisible technicians » who are in charge of managing scientific data in laboratories. The paper shows how the development of a large-scale database project goes hand in hand with processes of establishing different degrees of visibility for information managers and their work. It discusses issues related to the invisibility of data documentation work, particularly this invisibility's impact on scientific knowledge processes. The invisibility of information managers appears to be related to a fundamental aspect of their work, « articulation work », characterized by activities of *bricolage*, translation and deletion. The invisible work of data management and documentation is that of endlessly redoing and reinventing.

Keywords: invisible work, invisible technician, information manager, metadata, standard, infrastructure

RESUMEN: LA CIENCIA EN RED. LOS GESTORES DE INFORMACIÓN “INVISIBLES” EN LA PRODUCCIÓN DE UNA BASE DE DATOS CIENTÍFICOS

El desarrollo tecnológico contemporáneo, específicamente de infraestructuras numéricas, permite observar nuevas formas de trabajo, incluso nuevas categorías de trabajadores, en el contexto científico. Basado en un estudio etnográfico de una red estadounidense de investigadores de ecología, este artículo se centra en el trabajo de los gestores de información (*information managers*), “técnicos invisibles” responsables de la gestión de datos científicos en los laboratorios. El estudio demuestra cómo el desarrollo de una gran base de datos viene acompañado de procesos de puesta en visibilidad e invisibilidad de los gestores de información y su trabajo. Además, examina los problemas de la puesta en invisibilidad de las actividades de documentación de datos en particular, sobre los procesos de producción de conocimientos científicos. La invisibilidad de los gestores de información parece vinculada a un aspecto fundamental de su trabajo, en este caso a un “trabajo de articulación” caracterizado por actividades de bricolaje, traducción y borrado. El trabajo de esas “pequeñas manos” que realizan la gestión y la documentación de datos en ciencias se rehace y reinventa constantemente.

Palabras claves: Trabajo invisible, técnico invisible, gestores de información, metadato, estándar, infraestructura