

Metadata, trajectoires et « énaction »

Florence Millerand¹ et Geoffrey C. Bowker²

Les progrès technologiques en matière de cyberinfrastructure (ou *e-science*) et l'explosion des données scientifiques dans de nombreux champs disciplinaires ont projeté sur le devant de la scène la question des enjeux liés à la conservation, à l'accès et au partage de l'information scientifique³. Il ne s'agit pas tant de la préservation et de l'échange de résultats de recherche (les publications), que des données dites *brutes* qui, une fois collectées, traitées, analysées, modélisées, permettent l'obtention de ces résultats de recherche. Ces enjeux se font particulièrement pressants dans les disciplines des sciences de l'environnement⁴ où les grandes questions scientifiques actuelles, la biodiversité ou le changement climatique par exemple, requièrent le passage à de nouvelles échelles spatiales et temporelles pour mener des recherches, d'une part comparatives et d'autres part, sur le long terme⁵. Si traditionnellement en écologie, les données étaient collectées par des chercheurs individuels ou en petits groupes, sur de petits espaces (inférieurs à un mètre carré) et sur de courtes périodes temporelles, les programmes de recherche actuels requièrent la manipulation de quantités de données bien supérieures à ce qu'un chercheur individuel, voire même un groupe de chercheurs réunis autour d'un projet, peuvent collecter et analyser⁶. Aussi les chercheurs sont-ils amenés à utiliser et réutiliser des jeux de données constitués par d'autres, dans des contextes disciplinaires différents et pour atteindre de nouveaux objectifs scientifiques.

¹ Laboratory of Comparative Human Cognition & Scripps Institution of Oceanography, University of California, San Diego.

² Center for Science, Technology and Society, Santa Clara University.

³ P. W. Arzberger, P. Schroeder, A. Beaulieu, G. C. Bowker, K. Casey, L. Laaksonen et al., « Promoting access to public research data for scientific, economic, and social development », *Data Science Journal*, 3 (29), 2004, pp. 135-152

⁴ J. F. Franklin, C. S. Bledsoe & J. T. Callahan, « Contributions of the Long-Term Ecological Research program », *BioScience*, 40 (7), 1990, pp. 509-515

⁵ G. C. Bowker, *Memory of practices in the sciences*, Cambridge, MIT Press, 2006.

⁶ W. K. Michener, J. W. Brunt, J. J. Helly, T. B. Kirchner & S. G. Stafford, « Nongeospatial metadata for the ecological sciences », *Ecological Applications*, 7 (1), 1997, pp. 330-342

L'un des défis majeurs qu'ont à relever les projets de développement d'infrastructure d'information scientifique (cyberinfrastructure) qui visent à favoriser la collaboration et le partage de données via les réseaux informatiques consiste précisément à assurer la bonne circulation des données à travers les diverses plates-formes techniques, environnements organisationnels, disciplines, institutions, etc., autrement dit à assurer l'interopérabilité des données. La principale difficulté tient au fait que la localisation, l'accès et l'évaluation de la validité des données sont extrêmement dépendants de la façon dont ces mêmes données auront été collectées, "étiquetées" et stockées. Si une simple note manuscrite rappelant le projet scientifique et les conditions d'expérimentation dans lesquelles un jeu de données a été constitué peut convenir à un échange entre deux collègues, il va sans dire qu'une documentation plus détaillée sera nécessaire dans le cas d'équipes pluridisciplinaires travaillant sur des études comparées et sur de longues échelles de temps. Dans ce dernier cas, les ensembles de données mobilisés pourront avoir été constitués dans le cadre de projets spécifiques – utilisant des protocoles dont les détails sont connus de leurs seuls auteurs – ou encore sur des données anciennes collectées et codées avec les instruments et terminologies de l'époque – dont les auteurs pourraient être passés à la postérité ! On comprend alors que de vives discussions sur le besoin de méthodes standardisées de gestion de l'information scientifique accompagnent les débats actuels sur les grands enjeux écologiques.

Envisagés le plus souvent comme de simples composants techniques, les standards développés pour assurer l'interconnexion des systèmes et, par là, la circulation des données jouent un rôle clef dans l'activité scientifique. La capacité de coopération des collectifs scientifiques, situés dans des contextes disciplinaires et institutionnels souvent très disparates, repose précisément sur l'existence d'infrastructures d'information et de collaboration communes, nécessairement fondées sur des standards partagés. Largement ignorés, invisibles (enchâssés dans une infrastructure technique prise comme une boîte noire), les standards constituent néanmoins les supports indispensables à la réalisation du travail cognitif distribué qui caractérise l'activité scientifique. Mieux comprendre leur fonctionnement, les dynamiques à l'œuvre dans leur adoption par les communautés scientifiques ainsi que les enjeux liés à leur développement constitue un

domaine de recherche extrêmement fécond pour qui s'intéresse aux modalités de la collaboration au sein des collectifs.

Ce texte porte sur un projet de développement d'une infrastructure d'information au sein d'une communauté de recherche en écologie, le réseau américain de recherches à long terme sur l'écologie (*Long Term Ecological Research network - LTER*)⁷, qui s'est engagé dans un processus de standardisation de la gestion de ses données scientifiques à travers l'adoption d'un standard de métadonnée⁸, baptisé *Ecological Metadata Standard* (ou EML). Ce processus de standardisation a débuté en 1996 avec le lancement du projet de développement d'un système d'information à l'échelle du réseau. Il s'est cristallisé en 2001 avec l'adoption du standard EML. Il fait, depuis⁹, l'objet d'une controverse au sein de la communauté que l'on peut résumer *grosso modo* en termes de « succès obtenu » *versus* « succès à venir ».

Ce sont précisément ces points de vue discordants qui nous intéressent ici et dont nous allons chercher à rendre compte à travers une analyse pragmatique et interactionniste faisant usage de méthodes ethnographiques. La question abordée n'est pas tant pas de savoir « à quoi tient » le succès de l'implantation d'un standard de gestion de l'information dans une organisation, mais plutôt à partir de quel *moment* et de quel *point de vue* le succès ou l'échec de cette implantation sont définis. Nous proposons de resituer ces récits en fonction des « places temporelles » qu'occupent leurs auteurs, c'est-à-dire dans le cours des différentes *trajectoires* à partir desquelles ils évoluent¹⁰. Plus précisément, nous rendons compte de ces évaluations en termes de succès obtenu ou de succès à venir en les traitant comme des résultats plus ou moins réussis *d'alignement* de différentes trajectoires. Enfin, nous proposons d'analyser cette dynamique de standardisation comme un processus *d'énaction*, pour mieux saisir l'enchâssement des dispositifs techniques dans les contextes organisationnels dans lesquels ils sont conçus, développés ou utilisés.

⁷ Le réseau français des Zones Ateliers du Centre national de la recherche scientifique poursuit des objectifs scientifiques similaires à ceux du réseau américain de recherches à long terme sur l'écologie (l'étude d'écosystèmes sur le long terme).

⁸ Les métadonnées sont « des données sur des données », c'est-à-dire des informations apportant des renseignements sur la nature de certaines autres données et qui permet ainsi une utilisation pertinente de ces dernières.

⁹ Le processus de standardisation était toujours en cours lors de la rédaction de cet article en 2005.

¹⁰ A. Strauss, *Continual permutations of action*, New York, Aldine de Gruyter, 1993.

1. Le développement d'infrastructures d'information à l'intersection de mondes sociaux

À l'image de l'activité scientifique, le développement d'infrastructures d'information pour les sciences (cyberinfrastructures) requiert la coopération de groupes d'acteurs variés et hétérogènes (chercheurs du domaine, spécialistes en technologie de l'information, chercheurs en informatique, organismes subventionnaires, etc.). L'étude des processus par lesquels une grande variété d'acteurs parvient à se coordonner et à coopérer a donné lieu à un vaste ensemble de travaux dans le domaine Science, Technologie, Société (ou *Science Studies*). Ces travaux ont proposé d'intéressantes conceptualisations de l'activité scientifique en termes de construction de réseaux sociotechniques¹¹ ou d'intersection de mondes sociaux¹². En mettant les processus de négociation au centre de l'analyse, ces travaux ont permis de mieux comprendre les dynamiques et jeux d'acteurs (*via* les notions d'intéressement, d'articulation, de traduction, d'objets frontières, etc.) par lesquels ces groupes diversifiés tiennent (ou non) ensemble. L'étude des processus par lesquels les faits sont construits – plutôt que l'étude des faits constitués, l'étude des processus par lesquels le travail est accompli – plutôt que l'étude des statuts et structures de division des tâches – sont au cœur de ces approches.

L'un des intérêts principaux de ces perspectives de recherche est de rendre compte des phénomènes en action, du point de vue des acteurs. Le temps est ici une « variable » et non un cadre à l'explication des phénomènes étudiés : « Le temps des innovations dépend de la géométrie des acteurs, pas du calendrier »¹³. Dans cette perspective, l'évolution d'un projet de développement technologique peut être comprise non pas en fonction d'un temps linéaire « objectif » (par exemple selon des phases d'émergence, de maturation, d'essoufflement, etc., dans une conception évolutionniste)

¹¹ M. Callon, « Éléments pour une sociologie de la traduction : la domestication des coquilles St-Jacques et des marins pêcheurs dans la baie de St. Brieuc », *L'Année Sociologique*, 36, 1986, pp. 169-208 ; M. Callon, « Actor network theory », in N. J. Smelser & P. B. Baltes (eds), *International encyclopedia of the social & behavioral sciences*, Amsterdam, New York, Elsevier, 2001, pp. 62-66 ; B. Latour, *Science in action : how to follow scientists and engineers through society*, Cambridge, Harvard University Press, 1987.

¹² J. H. Fujimura, « The molecular biological bandwagon in cancer research: where social worlds meet », *Social Problems*, 35 (3), 1988, pp. 261-283 ; S. L. Star & J. Griesemer, « Institutional ecology, 'translations,' and boundary objects : amateurs and professionals in Berkeley's museum of vertebrate zoology, 1907-1939 », *Social Studies of Science*, 19, 1989, pp. 387-420.

¹³ B. Latour, *Aramis ou l'amour des techniques*, Paris, La Découverte, 1993, p. 80.

mais plutôt en fonction des différentes temporalités des acteurs. Une telle perspective permet par ailleurs de mieux comprendre comment l'existence voire la réalité des projets ou des objets varie dans le temps, au fur et à mesure de l'engagement ou du retrait de groupes d'acteurs dans le développement de ces projets ou objets. Ainsi, bien qu'existant (par exemple sous la forme d'un prototype), un objet technique ne sera considéré « réel » qu'à partir du moment où certains groupes d'acteurs en particulier se seront ralliés au projet¹⁴.

La notion de trajectoire formulée dans le cadre de la théorie des mondes sociaux¹⁵ nous semble particulièrement intéressante pour comprendre le caractère temporel et processuel d'un phénomène en évolution. Cette notion renvoie d'une part au cours d'un phénomène, donc à son évolution dans le temps, et d'autre part aux processus de gestion de ce phénomène et de ses conséquences, c'est-à-dire aux actions et aux interactions qui contribuent à son évolution¹⁶. Ainsi, la trajectoire d'un projet de développement technologique renverra aussi bien au déroulement du projet lui-même qu'à l'organisation du travail qui y est associée et aux conséquences des activités des acteurs sur le projet et sur leurs actions futures.

Dans cette perspective, chaque acteur évolue à partir d'une trajectoire qui correspond à un passé, un présent et un futur envisagé. Ainsi, un événement tel qu'un bogue informatique pourra constituer un problème majeur pour un utilisateur lorsqu'il constituera une tâche supplémentaire dans la routine quotidienne du réparateur. Ce même bogue s'inscrira dans une suite d'événements passés et impliquera des actions futures distinctes chez l'utilisateur et le réparateur : un historique d'installations de programmes instables et de négligences en matière d'habitudes de téléchargement chez l'utilisateur *versus* le énième cas de réinitialisation de systèmes dans la journée pour le réparateur ; l'installation d'un logiciel anti-virus chez l'utilisateur *versus* le perfectionnement d'une routine de débogage chez le réparateur ; etc.

L'ensemble des actions liées à une trajectoire peut engager de multiples acteurs, eux-mêmes habitant différents mondes sociaux et suivant de multiples trajectoires,

¹⁴ B. Latour, op. cit.

¹⁵ A. Strauss, op. cit.

¹⁶ A. Strauss, op. cit, pp 52-54.

chacun ayant sa propre vision du déroulement du phénomène et sa propre conception des actions nécessaires à sa gestion et à sa mise en forme. Les dispositifs techniques suivent également des trajectoires¹⁷. Un logiciel informatique a une histoire (une idée, un algorithme, une version bêta, etc.), un cycle de vie, un futur envisagé. Ainsi, un bogue informatique pourra constituer un point de non-retour en obligeant la révision partielle voire l'abandon du logiciel ou alors ne constituer qu'un événement de routine dans son cycle de vie.

Toute trajectoire étant elle-même composée de l'interaction de différentes trajectoires, il se produit nécessairement un travail d'alignement (ou de désalignement) de ces trajectoires, via l'articulation (ou la désarticulation) des perspectives et des actions. Ainsi, lorsqu'une trajectoire change de forme, des trajectoires croisées peuvent se réaligner, par exemple pour maintenir la trajectoire dans une direction souhaitée, alors que d'autres pourront être écartées¹⁸. Ce « réglage » mutuel des trajectoires les unes par rapport aux autres¹⁹ donne à observer des phénomènes d'ajustements dynamiques voire de transformation des personnes, des objets, ou des projets. On pourra alors observer l'altération identitaire d'un groupe d'acteurs ou la redéfinition du « script » d'un dispositif technique²⁰, ou encore l'ajustement d'une structure organisationnelle, l'émergence de nouveaux rôles, etc.

Ce sont justement ces points de contact entre les différentes trajectoires d'acteurs (humains et non humains) qui nous intéressent dans cette étude des processus de standardisation à l'œuvre au sein de la communauté de recherche LTER. Que se passe-t-il dans ce processus de standardisation des pratiques de gestion de données qui conduit un groupe d'acteurs à formuler une *autre* histoire sur le standard (une histoire que d'autres avaient déjà consignée dans les termes d'une *success story*) ? Quelles sont les trajectoires

¹⁷ S. Timmermans & M. Berg, « Standardization in action: achieving local universality through medical protocols », *Social Studies of Science*, 27 (2), 1997, pp. 273-305.

¹⁸ J. L. Lemke propose la notion d'hétérochronie (*heterochrony*) pour décrire les décalages entre les différentes échelles temporelles qui coexistent dans les milieux organisés. J. L. Lemke, « Across the Scales of Time : Artifacts, Activities, and Meanings in Ecosocial Systems », *Mind, Culture, and Activity*, 7 (4), 2000, pp. 273-290.

¹⁹ S. Timmermans, « Mutual tuning of multiple trajectories », *Symbolic Interaction*, 21 (4), 1998, pp. 425-440.

²⁰ M. Akrich, « The de-description of technical objects », in W. E. Bijker & J. Law (eds), *Shaping technology/building society : studies in sociotechnical change*, Cambridge, MIT Press, 1992, pp. 259-263.

qui s'entrecoupent et comment s'ajustent-elles ? Dans quelle mesure certaines trajectoires se trouvent-elles redéfinies ?

Dès lors, se pose la question de savoir quelle(s) trajectoire(s) suivre²¹. Car le choix de suivre une trajectoire en particulier plutôt qu'une autre peut conduire à l'adoption d'un nouveau regard sur la configuration sociale, technique et organisationnelle à l'étude. À qui bénéficie la configuration actuelle ? Aux dépens de qui ? Les travaux de Timmermans et Berg sur la construction sociale de protocoles médicaux ont révélé l'inégalité des pouvoirs et des capacités d'influence parmi les acteurs. Ils invitent à prêter attention à la manière dont les trajectoires sont modifiées (négociation, imposition ou coercition) de même qu'au travail souvent insidieux d'exclusion de trajectoires, réalisé par exemple au profit du maintien d'autres trajectoires²². De la même façon, les analyses féministes ont montré combien les analyses de réseaux pouvaient changer radicalement dès lors qu'on y intégrait la perspective de ceux qui avaient été enrôlés (et pas seulement de leurs porte-parole) ou encore de ceux qui avaient été exclus des traductions²³.

Plus généralement, les études d'infrastructures (*infrastructure studies*) ont permis de braquer le projecteur sur des phénomènes rarement étudiés, notamment le travail « invisible » réalisé en arrière-plan par des acteurs dont la performance est considérée d'autant meilleure que le résultat de leur travail demeure invisible²⁴. Le travail de maintenance et de soutien technique, qui n'apparaît qu'en cas de problèmes, en constitue un exemple éclairant²⁵. Dans cette perspective, l'étude du travail tel qu'il est accompli en pratique rend possible la mise au jour des éventuelles différences entre ceux qui réalisent

²¹ J. H. Fujimura, « On methods, ontologies, and representation in the sociology of science : where do we stand ? » in D. R. Maines (ed), *Social organization and social process*, New York, Aldine de Gruyter, 1991, pp. 207-248 ; S. L. Star, « 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, 1991, pp. 27-57 ; S. Timmermans, op. cit.

²² S. Timmermans, op. cit. ; S. Timmermans & M. Berg, op. cit.

²³ S. L. Star, op. cit.

²⁴ G. C. Bowker & S. L. Star, *Sorting things out : classification and its consequences*, Cambridge, MIT Press, 1999 ; S. L. Star & G. C. Bowker, « How to infrastructure », in L. A. Lievrouw & S. Livingstone (eds), *Handbook of new media. Social shaping and consequences of ICTs*, London, Thousand Oaks, New Delhi, Sage Publications, 2001, pp. 151-162 ; S. L. Star & K. Ruhleder, « Steps toward an ecology of infrastructure : design and access for large information spaces », *Informations Systems Research*, 7 (1), 1996, pp. 111-134.

²⁵ S. Shapin, « The Invisible Technician », *American Scientist*, 77, 1989, pp. 554-563.

le travail et ceux qui en sont récompensés²⁶. En choisissant de rendre compte de deux perspectives différentes sur le processus de standardisation au sein de la communauté LTER, nous cherchons délibérément à contribuer à la visibilité d'un point de vue en particulier, celui qui, multiple, diffus et mal articulé, raconte une histoire différente.

À travers cette étude, nous cherchons plus largement à nous interroger sur le « travail » que réalisent les dispositifs techniques, en particulier sur les plans sociaux et organisationnels. Berg a montré comment le dossier médical du patient « performait » le travail de soin médical à travers les pratiques d'écriture et de lecture qu'il impliquait²⁷. Supports par lesquels s'articulent les pratiques de travail²⁸, infrastructures organisationnelles²⁹, réservoirs d'inscriptions³⁰, distributeurs de cognition³¹, nous appréhendons les dispositifs techniques comme constitutifs des activités liées au travail.

Nous appréhendons le standard EML en tant que support à la coordination de différents mondes sociaux (chercheurs du domaine, développeurs du standard, gestionnaires d'informations chargés de l'implémentation de ce dernier, etc.). *A priori* perçus comme un ensemble de problèmes techniques dont le défi majeur consiste à développer « le » bon outil qui en facilitera la résolution, ce standard technique et le processus d'implémentation associé apparaissent en réalité être intimement imbriqués dans la structure du travail scientifique. Tant le standard EML que le projet de système d'information dans lequel il s'inscrit incarnent en effet certaines conceptions sur l'organisation du travail, c'est-à-dire certaines configurations d'acteurs, d'outils et de données scientifiques³².

²⁶ S. L. Star & A. Strauss, « Layers of silence, arenas of voice : the ecology of visible and invisible work », *JCSW*, 8, 1999, pp. 9-30.

²⁷ M. Berg, « Practices of reading and writing: the constitutive role of the patient record in medical work », *Sociology of Health & Illness*, 18 (4), 1996, pp. 499-524.

²⁸ S. Timmermans & M. Berg, *The gold standard : a sociological exploration of evidence-based medicine and standardization in health care*, Philadelphia, Temple University Press, 2003.

²⁹ M. Berg & G. C. Bowker, « The multiple bodies of the medical record: toward a sociology of an artifact », *The Sociological Quarterly*, 38 (3), 1997, pp. 513-537.

³⁰ E. Hutchins, *Cognition in the wild*, Cambridge, MIT Press, 1995 ; B. Latour, op. cit.

³¹ F. Millerand, D. Ribes, K. S. Baker & G. C. Bowker, « Comparative interoperability project: collaborative science, interoperability strategies, and distributing cognition », communication au Workshop Artifacts & Collectives, Lyon, 2005.

³² D. Ribes, K. S. Baker, F. Millerand & G. C. Bowker, « Comparative interoperability project : configurations of community, technology, and organization » in *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries*, New York, ACM Press, 2005, pp. 65-66.

Enfin, nous faisons appel à une conceptualisation des processus de changement organisationnel associés au changement technologique en termes d'« éniation » de technologies. Dans la lignée du concept d'*enactment* proposé par Weick³³, Fountain³⁴ nous invite à distinguer une technologie « objective », c'est-à-dire un ensemble de composants techniques matériels et logiciels (Internet par exemple) d'une technologie « éniatée », c'est-à-dire le dispositif technique tel qu'il est perçu, conçu ou utilisé, en pratique, dans un contexte particulier. Dans cette perspective, la façon dont les acteurs « éniatent » les dispositifs techniques est directement dépendante de leur enchiâssement dans des structures cognitives, culturelles, sociales et institutionnelles. Les agencements organisationnels (caractérisés par des routines, des normes, des valeurs, des politiques, etc.) médiatisent ainsi l'éniation des technologies qui, en retour, contribue à refaçonner ces agencements organisationnels.

Nous proposons donc d'examiner « en action » la composition d'histoires – au pluriel – sur un même processus de standardisation, en resituant les artefacts, les acteurs et les récits dans les contextes qui les ont fait émerger. Dans une démarche d'analyse ethnographique³⁵, nous cherchons à rendre compte du travail de négociation par lequel certaines configurations sociotechniques sont remises en jeu alors que de nouvelles sont proposées³⁶. Une telle perspective devrait nous permettre de mieux comprendre les dynamiques sociales et organisationnelles à l'œuvre dans les projets de développement d'infrastructures technologiques de grande envergure.

2. La communauté de recherche LTER et le standard EML (*Ecological Metadata Language*)

Le programme scientifique américain de recherches à long terme sur l'écologie (*Long Term Ecological Research program - LTER*) se présente sous la forme d'un réseau de plus de 1200 chercheurs et étudiants. Lancé en 1980, le réseau regroupe 26 sites (ou

³³ K. E. Weick, *The social psychology of organizing*, Addison-Wesley, 1979.

³⁴ J. E. Fountain, *Building the virtual state : information technology and institutional change*, Washington, Brookings Institution Press, 2001.

³⁵ L'étude est réalisée à partir de notes d'observation *in situ*, d'entretiens et d'analyses documentaires (littérature scientifique et technique produite au sein de la communauté de recherche étudiée).

³⁶ H. Karasti, K. S. Baker & G. C. Bowker, « Ecological storytelling and collaborative scientific activities », *SIGGROUP Bulletin*, 23 (2), 2002, pp. 29-30.

stations de recherche) constitués autour de l'étude d'un biome (par exemple un désert chaud, un estuaire côtier, une forêt coniférienne tempérée, une toundra arctique, etc.) répartis dans la zone continentale des Etats-Unis et de l'Antarctique, auxquels s'ajoute un 27^e site chargé de l'administration et la coordination de l'ensemble. Ce programme scientifique vise essentiellement l'amélioration des connaissances sur les changements environnementaux *via* la collaboration interdisciplinaire et la conduite de projets de recherche sur le long terme.

La conduite d'études à long terme sur les écosystèmes constitue la particularité première de ce programme de recherche hors du commun. À titre comparatif, alors que la plupart des recherches sur l'écologie couvrent des périodes de temps allant de l'heure à la journée, les études réalisées au sein du programme LTER peuvent porter sur des mois voire des siècles. De la même façon, alors que la station (*plot*), cet échantillon d'écosystème rarement supérieur à un mètre carré, constitue l'échelle de référence des recherches environnementales, les recherches comparatives à l'échelle régionale, continentale ou globale sont celles privilégiées au sein du programme LTER.

Les enjeux sont de taille. Alors que les cycles de financement de la recherche scientifique consistent généralement en des programmes de 5 ans, que les carrières des chercheurs s'étendent sur des périodes moyennes de 30 ans, les changements dans les écosystèmes suivent quant à eux des cycles variant entre une cinquantaine et une centaine d'années. Faire coexister ces différentes échelles temporelles constitue le défi majeur du programme LTER.

Si la préservation des données dans le temps, leur conservation dans des conditions appropriées de façon à garantir leur réutilisation adéquate, à la fois dans le présent et dans le futur a toujours été une priorité au sein des différents sites du réseau LTER, ces enjeux ont pris une nouvelle ampleur avec le projet de développement d'une infrastructure d'information visant à encourager le partage des données au sein de la communauté.

Chacun des 26 sites du réseau assure la gestion des données de recherche produites localement et dispose généralement de son propre système d'information (ses propres bases de données). Un gestionnaire d'information assume la responsabilité du travail de conception et de maintien des infrastructures locales. À l'échelle du réseau, les

données se trouvent donc stockées de manière autonome et dispersées à travers les sites, ce qui rend la tâche de localisation et d'accès à ces données relativement complexe et laborieuse. Le projet d'une infrastructure d'information en réseau fondée sur un système de bases de données fédérées a alors été formulé pour faciliter l'accès et le transfert des données grâce à la création d'un système global qui soit interopérable avec les systèmes locaux.

Ce projet, lancé en 1996, fait face à trois défis majeurs : l'hétérogénéité des données qui circulent au sein de la communauté de recherche, leur grande dispersion et l'extrême variété des systèmes de stockage et de catalogage³⁷. Au-delà de la diversité des données liées au projet scientifique à l'étude, l'organisation et les formats de données peuvent être extrêmement disparates selon les conditions et protocoles de collecte utilisés. À titre d'exemple, des données de mesure des quantités de chlorophylle prélevées ponctuellement en mer seront organisées dans des fichiers autonomes correspondants au nombre de croisières effectuées par année alors que les mêmes mesures prélevées à l'année dans un lac seront conservées dans un seul et même fichier. Par ailleurs, les systèmes de catalogage locaux reposent généralement sur une méta information (ou métadonnée) qui n'est pas forcément compréhensible en dehors de la discipline, du projet de recherche ou encore du site à l'intérieur duquel ces systèmes ont été constitués³⁸. Ainsi, des unités de mesure « spéciales » (personnalisées) peuvent être créées pour les besoins d'analyse d'un projet de recherche particulier (par exemple l'unité « nombre-de-feuilles-par-partie-supérieure-de-pousse »).

Dans ce contexte, le besoin de standards en matière de gestion des données s'est rapidement imposé au sein du réseau LTER. Plus précisément, la nécessité de développer des méthodes standardisées de description des métadonnées sur les jeux de données est apparue incontournable.

Idéalement, les métadonnées (données sur les données) comprennent tous les détails nécessaires à un usage secondaire ultérieur d'un jeu de données. Ainsi, elles comprendront des informations aussi détaillées et diversifiées que les noms des chercheurs auteurs des données, le titre du projet pour lequel ces données ont été

³⁷ M. B. Jones, C. Berkley, J. Bojilova & M. Schildhauer, « Managing scientific metadata », *IEEE Internet Computing*, 5 (5), 2001, pp. 59-68.

³⁸ G. C. Bowker, « Biodiversity Datadiversity », *Social Studies of Science*, 30 (5), 2001, pp. 643-684.

collectées, le résumé du projet, des mots-clefs, le type de biome, les techniques d'échantillonnage, l'étalonnage de tel instrument et son numéro de série, les unités de mesure utilisées, etc.³⁹

Simple technicalités en apparence, les métadonnées revêtent une importance capitale dans la production scientifique dans la mesure où elles conditionnent l'accès aux données, en garantissent l'intégrité et en délimitent les cadres d'interprétation. Dans le cas d'études comparées ou de recherches sur le long terme, on peut s'attendre à ce que les équipements changent avec le temps, par exemple qu'ils gagnent en précision, d'où l'importance de documenter précisément l'étalonnage des instruments de mesure utilisés au moment de la collecte des données. De la même façon, la possibilité de réaliser des analyses croisées sur des jeux de données physiques, biologiques ou chimiques portant sur une même zone géographique, dépendra de la qualité des métadonnées associées. En ce sens, les métadonnées assurent plus qu'un simple rôle d'étiquetage pour des fins de stockage et d'accès aux données. En permettant la création de nouveaux liens entre des données, autrement dit en autorisant la juxtaposition de certaines données autrement jamais rapprochées, elles peuvent participer à la formulation de nouvelles hypothèses de recherche et de résultats scientifiques.

Le langage de description de métadonnées en écologie EML (*Ecological Metadata Language*) constitue précisément un langage standardisé pour la rédaction de métadonnées dans le domaine particulier des sciences de l'environnement. C'est ce standard que la communauté de recherche LTER a décidé d'adopter lorsqu'elle s'est engagée dans un processus de standardisation de ses pratiques de gestion des données scientifiques.

3. Controverses

Le processus de standardisation dans lequel la communauté LTER s'est engagée poursuit deux grands objectifs scientifiques : 1) la promotion de la collaboration interdisciplinaire *via* le partage facilité de données et 2) l'amélioration de la conservation des données sur de longues périodes de temps⁴⁰. Si l'ensemble des acteurs semblent s'être

³⁹ W. K. Michener et al, op. cit.

⁴⁰ J. E. Hobbie, S. R. Carpenter, N. B. Grimm, J. R. Gosz & T. R. Seastedt, « The US Long Term

entendus sur ces grands objectifs et sur le choix du standard EML, des voix discordantes se sont élevées au moment de la mise en œuvre effective du standard.

Nous rendons compte ici de deux discours, portés par deux catégories d'acteurs, qui racontent des histoires différentes sur le processus de standardisation et sur le standard EML : l'une *plutôt* réussie et l'autre *pas encore* réussie. Le premier récit présente le point de vue de ceux que nous appelons les « développeurs » du standard, ce qui inclut l'équipe d'experts ayant développé EML ainsi que les coordinateurs du réseau LTER. Le deuxième récit présente le point de vue de ceux que nous appelons les « énéacteurs » du standard, c'est-à-dire les gestionnaires d'information chargés de sa mise en œuvre au sein des sites. Au moment de l'étude, le premier récit dominait clairement le deuxième, il était déjà formalisé (si l'on peut dire) et même consigné dans des rapports, à la différence du deuxième qui, diffus et non écrit, était plus difficile à circonscrire.

3.1 Récit n°1: « EML est un succès : l'ensemble de la communauté LTER l'a adopté »

La version 1.0 d'EML naît en 1997 au sein du *National Center for Ecological Analysis and Synthesis* (NCEAS) à l'Université de Californie à Santa Barbara, avec le travail d'un chercheur en informatique appliquée à l'écologie, aidé par deux étudiants de troisième cycle. EML répond avant tout aux préoccupations internes du centre NCEAS, qui réalise, dès sa création en 1995, combien l'absence d'un système uniformisé de catalogage des données scientifiques limite sa capacité à mener des activités d'analyse et de synthèse dans le domaine, notamment la conduite de recherches comparatives. Une demande de subvention pour le développement d'un standard de métadonnées en écologie est alors déposée à la *National Science Foundation*. Elle est accordée en 1997, et le développement d'EML commence.

Sur le plan technique, le standard repose sur deux standards émergents (SGML et XML). Sur le plan du contenu, il reprend les grands types de descripteurs de données (métadonnées) en usage dans le domaine, tels que reconnus par l'*American Society of Ecology*. Les versions 1.0 à 1.4 d'EML sont développées entre 1997 et 1999. Elles sont mises à l'épreuve au sein du centre NCEAS. Compte tenu de la nature des problèmes

rencontrés, on envisage de réviser entièrement l'approche adoptée. Une deuxième demande de subvention est déposée et obtenue, soutenue cette fois par plusieurs partenaires dont la communauté de recherche LTER. L'équipe de développement jusque-là composée de trois personnes se transforme en collaboratoire⁴¹, sorte de plateforme collaborative fondée sur la participation volontaire et ouverte à l'ensemble de la communauté des sciences de l'environnement. Ce modèle de développement ouvert ne recueille pas immédiatement le succès envisagé, même si l'équipe accueille quelques nouveaux développeurs dont, pour la première fois, un gestionnaire d'information du réseau LTER. Le développement d'EML se poursuit, d'importants changements structurels sont apportés. Dix-sept versions d'EML seront produites entre 1999 et 2002.

En 2001, l'équipe estime avoir en main une version stable d'EML (version 1.9 qui deviendra EML 2.0 bêta). On décide de la présenter au congrès annuel des gestionnaires d'information du réseau LTER qui se tient à Madison aux Etats-Unis. Les réactions sont extrêmement positives, les gestionnaires d'information reconnaissent l'utilité d'un tel standard pour la communauté LTER et décident *d'adopter* EML.

La version 2.0 est lancée dans la foulée, le réseau LTER (la plus importante communauté de recherche en sciences de l'environnement) a adopté le standard EML, le projet EML est un succès.

3.2 Récit n°2: « EML n'est pas (encore) un succès : il faut (d'abord) le re-développer pour pouvoir l'exploiter »

Au moment de la création d'EML en 1997, les 26 sites du réseau LTER disposent déjà de systèmes de catalogage pour gérer leurs données scientifiques. Ces systèmes sont plus ou moins formalisés selon les sites, c'est-à-dire qu'ils n'utilisent pas nécessairement de vocabulaires standardisés, même si certains d'entre eux reprennent *grosso modo* les descripteurs de données reconnus par l'*American Society of Ecology*. En 1996, le lancement du projet de système d'information à l'échelle du réseau stimule les

⁴¹ G. M. Olson et al., « Technology to support distributed team science : The first phase of the Upper Atmospheric Research Collaboratory (UARC) », in G. M. Olson, T. Malone & J. Smith (eds), *Coordination Theory and Collaboration Technology*, Hillsdale, Lawrence Erlbaum Associates, 2001, pp. 761-783.

discussions en matière de procédures standardisées de gestion de données et encourage le développement d'outils chez l'ensemble des gestionnaires d'information de la communauté. Cependant, aucune initiative à l'échelle du réseau n'est encore développée. En 2001, le projet EML est favorablement accueilli par les gestionnaires d'information, qui décident de façon consensuelle de l'adopter. L'implémentation commence.

Si quelques sites commencent le travail d'implémentation du standard relativement rapidement, la majorité d'entre eux se heurtent à d'importantes difficultés : le standard est complexe et il est difficile d'en comprendre tous les aspects ; les outils techniques sensés faciliter l'implémentation du standard s'avèrent inutilisables ; l'ampleur du travail à accomplir est considérable et les ressources à disposition extrêmement limitées, certains sites doivent s'engager dans une restructuration complète de leurs pratiques de gestion de données.

De nombreuses solutions *ad hoc* voient le jour, par exemple des outils « maisons » que les gestionnaires d'information s'échangent pour faciliter le travail de conversion des systèmes locaux vers le format requis par le standard EML.

Les gestionnaires d'information organisent deux ateliers consacrés à l'implémentation d'EML en 2003 et 2004, qui débouchent sur la rédaction d'un document de synthèse sur les « meilleures pratiques » en matière d'implémentation d'EML. La planification et la réalisation du travail s'en trouvent grandement facilitées sur de nombreux sites. Suivra un plan d'implémentation en cinq étapes, formulé conjointement par les gestionnaires d'information et les coordinateurs du réseau LTER.

En août 2005, au congrès annuel des gestionnaires d'information à Montréal, le bilan est mitigé, l'implémentation d'EML est un processus complexe et laborieux dont les retombées en termes d'amélioration des processus de gestion des données restent difficiles à identifier. EML n'est pas encore un succès, il faut d'abord le re-développer partiellement pour pouvoir l'exploiter.

4. Trajectoires

D'un côté, nous avons une histoire à succès (*success story*) sur le standard EML à la mesure de son adoption par la communauté de recherche LTER. De l'autre, nous avons une histoire mi-figue mi-raisin à la mesure des conditions de son développement au sein

des sites. En 2001, alors que la première se présente comme une histoire quasiment close, la deuxième ne fait que commencer.

Une lecture simpliste de ces deux récits consisterait à dire que la mesure du succès du processus de standardisation diffère en fonction des différentes phases à partir desquelles on se situe (ici les phases de conception et de développement du standard *versus* celles de déploiement et d'implémentation). En d'autres termes, on comprendrait que les gestionnaires d'information ne puissent reconnaître la réussite du projet à *ce stade* – sous-entendu, ils la reconnaîtront de toute façon une fois le projet terminé. Cette conception évolutionniste des projets de développement technologique qui évolueraient en fonction d'un temps objectif ne nous apprend cependant rien sur ce qui se passe réellement durant ces périodes d'émergence, de développement, de maturation, d'implémentation, etc. En outre, une telle perspective reviendrait à privilégier le premier point de vue sur le deuxième, en considérant le succès du standard comme étant *donné* d'avance – que le temps qui passe viendrait naturellement confirmer. Ce serait, par ailleurs, privilégier l'invention (le standard) sur l'innovation (sa mise en pratique)⁴².

Il nous semble qu'une analyse temporelle des projets de développement technique aurait plutôt intérêt à chercher à comprendre leur évolution en fonction des différentes temporalités dans lesquelles ils s'intègrent. Il devient alors possible de rendre compte, *du point de vue* des acteurs, de l'ensemble des événements – incluant les périodes plus troubles dont certains rechignent à parler tandis que d'autres cherchent leurs mots.

4.1 Trajectoires multiples

Il est intéressant de préciser que l'ensemble des acteurs concernés par ce processus de standardisation (développeurs du standard EML, coordinateurs du réseau LTER, gestionnaires d'information, chercheurs du domaine...) ont adhéré – et continuent d'adhérer – au « projet EML », c'est-à-dire à l'idée d'un standard de métadonnées qui permettrait l'échange et le partage de données à l'échelle du réseau LTER et au-delà. En ce sens, il ne s'agit pas d'un cas d'imposition d'un standard par un groupe d'acteurs (les développeurs et les coordinateurs) à un autre (les gestionnaires d'information) qui y serait

⁴² J. Schumpeter, *The theory of economic development*, Cambridge, Harvard University Press, 1934.

hostile et résistant. Ces derniers se sont d'ailleurs toujours montrés extrêmement favorables au projet⁴³. C'est au moment de la mise en œuvre effective du standard sur les sites – alors qu'un vaste ensemble de problèmes émerge – que les premières voix discordantes se font entendre.

La mise au jour de ces difficultés et la controverse qui a suivi ont contribué à remettre en cause le statut même d'EML en tant que standard. Deux ans après son adoption, lorsqu'une enquête révèle qu'EML n'a pas encore fait l'objet d'une implémentation complète dans un seul des 26 sites du réseau et que les outils développés expressément pour la réalisation de cette tâche restent inutilisés, EML semble bien n'avoir le titre de standard que sur le papier.

La juxtaposition des deux récits révèle la confrontation de deux visions sur le standard EML. Si en 2001, EML existe en tant que standard pour l'équipe d'experts qui l'a développé, il n'a pas encore ce statut pour les gestionnaires d'information qui doivent d'abord le retravailler et l'adapter pour pouvoir l'utiliser. Ce dialogue imaginaire inspiré de l'histoire du projet Aramis racontée par Latour⁴⁴ représente le clivage entre ces deux groupes d'acteurs : "EML 2.0 existe, le gros est fait, il ne reste qu'à implémenter" (les développeurs). "Il ne reste qu'à ? Mais un standard de métadonnées, ce n'est pas qu'un langage, aussi perfectionné soit-il, ce n'est pas qu'une implémentation, c'est aussi et avant tout une *intégration* de données" (les gestionnaires d'information). En d'autres termes, en 2001, le standard EML est un standard de métadonnées... sans données.

Nous proposons de comprendre ces différentes perspectives sur le standard EML en les situant dans les différentes trajectoires des acteurs.

Ainsi, du point de vue de ses développeurs, le standard EML est avant tout un *projet de recherche et développement*. Ce projet vise la création d'un langage standardisé de description de métadonnées à l'avant-garde des réalisations en la matière et nourrit l'ambition d'en faire le standard de référence en sciences de l'environnement. Dans cette perspective, le développement du standard et son adoption par la communauté de

⁴³ K. S. Baker & H. Karasti, *The long term information management trajectory : working to support data, science and technology*, San Diego, SIO Report, 2005 ; H. Karasti & K. S. Baker, « Infrastructuring for the long-term : ecological information management », in *Proceedings of the Hawaii international conference on system sciences (HICSS'37)*, Hawaii, Computer Society Press (CD-ROM), 2004, 10 p.

⁴⁴ B. Latour, op. cit.

recherche la plus vaste qui soit en science de l'environnement constituent les principaux critères de succès du projet.

Du point de vue des coordinateurs du réseau LTER, le standard EML représente la *clef de voûte de la stratégie d'interopérabilité des données* du projet d'infrastructure d'information à l'échelle du réseau, c'est-à-dire la condition de la bonne circulation des données scientifiques à travers les différentes plates-formes techniques, environnements organisationnels et disciplinaires au sein de la communauté LTER. Aussi, le succès du projet est à la mesure de l'adoption et de l'utilisation du standard EML à l'échelle du réseau.

Dans la perspective des gestionnaires d'information, le standard EML représente un *ensemble d'outils et de pratiques pour gérer plus efficacement les données scientifiques*, notamment en améliorant la qualité des métadonnées produites au sein de chacun des sites. En cela, l'incorporation réussie de ce nouvel outil – et des nouvelles façons de faire qui l'accompagnent – au sein des infrastructures sociotechniques locales constitue le critère majeur du succès du projet EML.

Enfin, du point de vue des scientifiques appartenant à la communauté de recherche LTER, le standard EML est un dispositif technique qui permettra la réalisation de recherches multi-sites grâce à un meilleur accès aux données et un meilleur partage de ces dernières ; il permettra également une meilleure diffusion des données au-delà du réseau LTER. De ce point de vue, la capacité à réaliser des recherches croisées sur de vastes ensembles de données *via* une interface unique constitue le critère principal de réussite du projet.

4.2 Alignement de trajectoires

Nous pouvons appréhender le travail d'implémentation réalisé par les gestionnaires d'information comme un processus d'appropriation⁴⁵ du standard au cours duquel se réalise un travail d'alignement de trajectoires. L'appropriation du langage EML sur les différents sites se traduit en effet par l'ajustement du dispositif technique aux

⁴⁵ F. Millerand, « Usages des NTIC, les approches de la diffusion, de l'innovation et de l'appropriation (1re partie) », *COMMposite*, 98 (1), 1998, http://commposite.org/v1/98.1/articles/ntic_1.htm ; F. Millerand, « Usages des NTIC, les approches de la diffusion, de l'innovation et de l'appropriation (2e partie) », *COMMposite*, 99 (1), 1999, http://commposite.org/v1/99.1/articles/ntic_2.htm

contextes locaux et par l'adaptation des pratiques préexistantes à de nouvelles façons de faire. Concrètement, on observe un véritable travail de bricolage chez les gestionnaires d'information qui cherchent à incorporer au standard (générique) la contextualité (locale) qui lui donnera sa signification et qui permettra son utilisation.

La trajectoire d'EML qui, selon le premier récit naissait des grands descripteurs de données écologiques en usage dans le domaine, devenait un projet de recherche-développement du centre NCEAS, puis le standard de métadonnées de la communauté LTER pour, enfin, devenir le standard de métadonnées de référence en sciences de l'environnement, semble faire l'objet d'une certaine réorientation au moment de son passage par le réseau LTER.

À partir de ce moment en effet, le découpage et l'ordonnement du projet EML en phases de conception, développement, déploiement et implémentation se trouvent bouleversés. On fait du re-développement en phase d'implémentation, ce qui conduit à revoir les principes de conception, puis à re-développer pour ré-implémentation, etc. Le projet EML se modifie, l'ensemble des trajectoires doivent se réaligner.

Que se passe-t-il donc lors de cette phase d'implémentation du standard EML qui nécessite plus que... du travail d'implémentation ?

Spontanément, les gestionnaires d'information évoquent le manque d'outils appropriés pour convertir les systèmes de métadonnées locaux en « format » EML. Ils se plaignent également d'une mauvaise compréhension des processus réels d'implémentation chez les coordinateurs du réseau dont les attentes leur apparaissent irréalistes. C'est que les pratiques de gestion de données ne sont pas seulement dépendantes des types d'infrastructures techniques, elles sont aussi et surtout intimement liées à la nature des projets de recherche étudiés, aux cultures disciplinaires et organisationnelles des sites, bref à la structure locale du travail scientifique.

Les deux extraits suivants illustrent d'une part le caractère local et contingent du travail scientifique, et d'autre part la complexité de la tâche des gestionnaires d'information qui sont chargés du catalogage des données de recherche. Le premier extrait présente un exemple d'usage impropre d'une unité de mesure, fondé sur des conventions partagées au sein d'un environnement de travail, en l'occurrence un site du réseau LTER. Le deuxième extrait présente des exemples d'unités de mesure utilisées par

les chercheurs du réseau LTER qui sont *a priori* identiques mais qui peuvent signifier des choses différentes en fonction des environnements disciplinaires dans lesquels elles peuvent être utilisées.

(1) J'avais reçu des données sur les substances nutritives et mes unités arrivaient en micromoles, avec le symbole μ pour micron et la lettre majuscule M, micromoles. Quand j'ai commencé à me lancer dans EML, j'ai dû chercher à comprendre, bon, qu'est-ce que c'est que cette unité exactement. Et en creusant plus profondément, en allant au lab. où les données sont traitées, j'ai réalisé que ce n'était pas des micromoles mais des micromoles *par litre*. Et bon, comme je ne suis pas chimiste, ça ne me disait absolument rien, moi je m'occupe d'organiser et de mettre en ligne ces données, point. Mais ça m'a vraiment ouvert les yeux sur le fait que j'avais là un problème beaucoup plus important que celui que j'imaginai : j'ai des personnes ici qui formulent des choses en micromoles alors que c'est faux. Mais voilà, c'est comme ça qu'on travaille, qu'on partage l'information et personne ne remet ça en question. C'est plutôt intéressant. J'ai donc commencé, fichiers après fichiers, à essayer de tout adapter en EML. Je me retrouve avec cette liste d'unités spéciales que je compile, et bon j'essaie de faire de mon mieux pour les décortiquer... Là je vais voir soit mon patron soit le collaborateur qui m'a donné ces données, et là je dois m'asseoir avec eux et leur demander : est-ce que vous pouvez vérifier s'il vous plaît, si vous aviez à décrire cette unité en EML, est-ce que c'est correct comme ça, est-ce que vous utilisez cette unité de la bonne façon, est-ce que vous utilisez le nom que tout le monde utilise, ce genre de choses... (IM_L)

(2) Par exemple, pour mesurer la concentration, on a « micromoles par litre » et « micromolar ». Techniquement, les deux sont des micromoles par litre et donc sont équivalentes en grandeur. Mais leur portée est différente car les micromoles par litre peuvent être utilisées pour un composant dissolu ou particulaire alors que micromolar s'utilise uniquement pour un composant dissolu. Donc elles ne sont pas exactement interchangeables...

On a aussi « micromoles par litre » et « millimoles par mètre cube ». Ces unités sont équivalentes en grandeur, mais différentes disciplines ont des préférences pour l'une ou l'autre.

Par ailleurs, si vous êtes en mer, vous pourrez fonctionner en micromoles par kilogramme et en micromoles par mètre cube, seulement si vous restez au niveau de la mer, parce que la conversion de l'une à l'autre dépend de la pression... (IM_M)

4.3 L'exemple du dictionnaire en tant que stratégie d'articulation

Au début de l'année 2005, alors que les outils créés par les développeurs ne sont toujours pas utilisés par les gestionnaires d'information et que l'implémentation traîne en longueur, les coordinateurs du réseau lancent le développement d'un nouvel outil sensé accélérer l'implémentation du standard sur les sites. Parallèlement et en partie en réaction à ce projet, quelques gestionnaires d'information initient le développement d'un outil « maison » : un répertoire des unités de mesure.

L'une des difficultés principales auxquelles les gestionnaires d'information font face est liée à la complexité du travail de traduction de leurs métadonnées en langage EML, notamment en ce qui concerne les unités de mesure. D'une part, le dictionnaire des unités de mesure fourni avec le standard EML répertorie essentiellement des unités de mesure physiques des phénomènes écologiques – alors que la majorité des sites du réseau LTER travaillent à partir d'unités de mesure écologiques. D'autre part, il est extrêmement difficile de décrire, dans un langage standardisé, les unités de mesure « spéciales » ou personnalisées, c'est-à-dire les unités formées expressément pour un projet de recherche et qui n'ont de sens qu'à l'intérieur de ce projet de recherche.

Aux prises avec ces difficultés, quelques gestionnaires d'information commencent alors à s'échanger les listes d'unités de mesure (incluant les unités spéciales) qui sont utilisées dans leur site, afin de comparer leurs traductions respectives et relever les éventuelles incongruités. Ces listes circulent sous la forme de feuilles de tableur via le courrier électronique. Rapidement, le projet de transformer ces listes en un répertoire des unités de mesure à l'échelle du réseau LTER est formulé. On envisage la réalisation d'un outil dynamique, accessible en ligne via l'Intranet LTER. L'équipe composée jusque-là uniquement de gestionnaires d'information intègre alors un membre de l'équipe des développeurs d'EML, par ailleurs également représentant du bureau de coordination du réseau. Un prototype intégrant les listes d'unités de six sites du réseau est développé. Il est présenté en août 2005 au congrès annuel des gestionnaires d'information de la communauté LTER, à la fois comme outil d'implémentation du standard EML et comme exemple de projet de collaboration réussie entre les gestionnaires d'information et les développeurs/coordonateurs.

Techniquement, cet outil permet à l'ensemble des gestionnaires d'information du réseau d'avoir accès à des définitions d'unités de mesure en langage EML (y compris certaines unités spéciales), de proposer des corrections au dictionnaire d'unités du standard, d'ajouter d'autres définitions d'unités, etc. Cependant, le rôle de ce répertoire d'unités de mesure dépasse considérablement cette simple fonctionnalité technique de conversion des formats de données. Il réalise avant tout un travail de coordination sociale.

Tant les initiatives *ad hoc* au niveau des sites (le partage d'outils maison ou de listes d'unités) que ce projet de répertoire à l'échelle du réseau peuvent en effet être compris comme des stratégies de ré-articulation du travail, d'une part entre les gestionnaires d'information eux-mêmes et d'autre part, entre eux et les développeurs/coordonateurs. Ce n'est pas seulement d'un outil technique de conversion dont les gestionnaires d'information avaient besoin, c'est avant tout d'un outil capable de soutenir tout ce travail de médiation informationnelle que l'énaction du standard a mis au jour. Comment décrire telle unité spéciale de mesure de la meilleure façon possible ? Peut-on considérer telle unité de mesure comme étant une unité de mesure « LTER » (reconnue à l'échelle du réseau) ou bien « spéciale » (propre à un site) ? Peut-on désormais ajouter telle unité de mesure au dictionnaire fourni avec le standard EML ?

En prenant en charge une partie de l'activité de traduction nécessaire à la communication entre des groupes d'acteurs distincts, le dictionnaire d'unités agit comme support à la coordination et à la coopération entre différents mondes. En d'autres termes, en créant un outil utilisable localement au niveau des sites qui puisse potentiellement contribuer à l'amélioration du standard, les gestionnaires d'information ont créé un objet frontière⁴⁶ apte à soutenir ce travail d'articulation entre énéacteurs et développeurs.

5. « Énaction »

Nous proposons de substituer aux notions techniques d'implantation et d'implémentation de standard la notion d' « énaction » pour mieux rendre compte de la complexité du travail réalisé par les gestionnaires d'information. Ceux-ci n'assurent pas seulement la mise en place du standard EML (son implantation) ou la programmation des quelques fonctionnalités supplémentaires qui permettront de le rendre fonctionnel (son implémentation), ils travaillent plutôt à son *interprétation*, voire à sa mise en actes (dans le sens théâtral du terme), ce qui implique d'adapter à la fois le standard et les pratiques de travail en place. Ces adaptations mutuelles peuvent alors être comprises non pas comme des activités de résistance, mais plutôt comme de nécessaires ajustements. Sans

⁴⁶ S. L. Star & J. Griesemer, op. cit.

ces réglages mutuels, le standard EML ne pourrait fonctionner en tant que standard au sein de la communauté LTER. C'est là toute la force de la notion d'énaction.

En quoi le standard EML contribue-t-il à des changements dans les mondes sociaux d'acteurs ? Comment, en même temps, ces mondes participent-ils à la transformation du standard ? Il nous semble que ces changements touchent, d'un côté aux identités d'acteurs et aux structures organisationnelles et de l'autre, au script du dispositif technique.

5.1 Nouveaux rôles et structures organisationnelles

A travers ce processus de standardisation, les gestionnaires d'information en tant que communauté de praticiens sont devenus plus « visibles » au sein du réseau LTER. De la même façon, certains aspects de leur activité, jusque-là peu connus, ont été rendus plus « explicites ». En 2004, alors que le développement d'outils maison fleurit au sein des sites d'une part, et que des gestionnaires d'information intègrent l'équipe de développeurs d'autre part, on reconnaît « officiellement » leur statut de développeurs à part entière (concrètement, en mentionnant leur contribution dans la documentation jointe au standard). Par ailleurs, le rôle actif qu'ils jouent dans l'activité scientifique se trouve en partie révélé par les initiatives locales de développement d'outils de conversion (le dictionnaire d'unités de mesure par exemple). La complexité de la tâche dont ils ont la charge, à savoir le travail de traduction de contenus scientifiques hérités de savoirs locaux dans un langage qui soit le plus « universel » possible (à l'échelle de la communauté scientifique en sciences de l'environnement), est ainsi mise en avant.

Cela étant dit, même si la transformation des modèles d'organisation au sein de la communauté LTER implique une réorganisation du travail (passage d'une structure de fonctionnement par sites à une infrastructure fédérée), et même si les rôles des gestionnaires d'information en ressortent considérablement transformés, leur statut de « techniciens » assurant un travail de soutien et de maintenance reste encore largement dominant dans les perceptions, notamment chez les chercheurs du domaine. En outre, même si l'équipe d'experts reconnaît leur statut de développeurs d'EML, l'évaluation de leur contribution reste ambiguë dans la mesure où les développeurs tendent à disqualifier

les initiatives des gestionnaires d'information, jugées trop « locales » et pas suffisamment « à la pointe de la technologie ». Ainsi le document de synthèse sur les « meilleures pratiques » produit par les gestionnaires d'information est-il jugé trop marqué par le contexte LTER pour pouvoir être intégré à la documentation du standard.

Il reste que l'ensemble des actions menées par les gestionnaires d'information dans le cadre de ce processus de standardisation ont permis de montrer qu'une autre configuration était possible, ne serait-ce que sur le plan très concret de l'allocation de ressources. Si le développement d'un standard de métadonnées requiert le type de financement associé à tout projet de recherche-développement, l'énaction d'un tel standard au sein d'une communauté de recherche comme LTER requiert elle aussi des moyens adaptés à l'ampleur du travail à réaliser.

Plus concrètement, les gestionnaires d'information ont contribué à la mise en place d'une nouvelle structure représentative, en l'occurrence un comité permanent formé à part égale de gestionnaires d'information et de chercheurs du domaine, dont la mission consiste à assurer un rôle de représentant et de conseiller en matière de développement de pratiques de gestion d'informations à l'échelle du réseau (le *Network Information System Advisory Committee* ou NISAC). Ce comité naît un an après l'adoption du standard EML, alors que les premières difficultés liées à son implémentation incitent les gestionnaires d'information à initier un dialogue avec les chercheurs du domaine. C'est par la voix de ce comité que le plan d'implémentation du standard en différentes étapes sera proposé puis adopté.

Outre cette nouvelle structure organisationnelle, une nouvelle forme de travail collaboratif à la jonction des niveaux locaux (sur les sites) et globaux (à l'échelle du réseau) s'est trouvée définie et expérimentée avec succès dans le cadre de ce processus de standardisation. Le projet de dictionnaire d'unités de mesure a constitué en effet une véritable innovation (par le bas) au sein de la communauté LTER dans la mesure où, d'une part il a permis la transformation d'une initiative locale en un projet de conception participative à l'échelle du réseau et d'autre part, il a créé un nouvel espace de collaboration entre deux groupes d'acteurs jamais associés aussi directement (gestionnaires d'information et développeurs/coordonateurs).

5.2 Redéfinition du standard

Si le standard lui-même a fait l'objet de multiples versions au cours de son développement, en partie sous l'impulsion des nouveaux membres qui intégraient l'équipe, c'est sous la forme d'une version achevée (une « boîte noire »), qu'il a été présenté en 2001 aux gestionnaires d'information du réseau LTER. Cela dit, cette boîte noire a été rouverte, et pas toujours du même côté, sur l'ensemble des sites du réseau.

Ainsi, on a identifié certaines lacunes du standard (en matière de définition d'unités de mesure par exemple). On a par ailleurs revu le plan d'implémentation prévu initialement en proposant un processus graduel à plusieurs niveaux, s'accommodant ainsi des rythmes d'intégration différenciés selon les sites. Plus largement, on a surtout révélé la nature des défis posés par l'énaction d'un standard à ce point générique qu'il pourrait, en théorie, s'appliquer à n'importe quel type de données⁴⁷.

Le projet EML comportait la définition d'un rôle pour les chercheurs, celui de décrire leurs jeux de données dans ce langage⁴⁸. En effet, les chercheurs ont toujours été présentés comme les futurs « usagers » du standard par ses développeurs. D'un côté, ils décrivent leurs ensembles de données en utilisant le langage EML, de l'autre ils peuvent réaliser des recherches croisées au sein de vastes bases de données grâce à ces descriptions standardisées. Il est intéressant de constater que le rôle de gestionnaire d'information n'a jamais été mentionné (du moins explicitement) dans les scénarios d'utilisation d'EML.

Et pourtant, en pratique, de nombreux chercheurs du réseau LTER ont refusé ce rôle, principalement par manque de temps et d'intérêt. Encore faut-il préciser que les métadonnées peuvent représenter, en quantité, l'équivalent du double des données qu'elles décrivent. Par ailleurs, l'ampleur de l'investissement requis pour l'apprentissage du langage EML (sans même parler des temps d'appropriation des outils de conversion) ont, chez la plupart des chercheurs, constitué un obstacle jamais surmonté. Ce sont les gestionnaires d'information qui, assez naturellement, ont repris ce rôle *a priori* dévolu à d'autres.

⁴⁷ M. Berg, *Rationalizing medical work : decision-support techniques and medical practices*, Cambridge, MIT Press, 1997.

⁴⁸ M. B. Jones et al, op. cit.

Certes, on peut saisir cette redistribution des rôles comme la marque d'une période de transition, et penser que les chercheurs du réseau LTER se « mettront » à EML au fur et à mesure que les outils se simplifieront et que le standard sera diffusé au sein des sciences de l'environnement. Cependant, la question de la formation des chercheurs à EML (et plus largement au nouveau contexte de gestion des données associé au projet d'une infrastructure intégrée) est actuellement en suspens, et il est fort probable que les gestionnaires d'information continueront d'assurer ce travail dans les prochaines années.

6. Conclusion

Pour conclure, d'un côté le standard EML ne change rien. Le partage des responsabilités reste le même (les gestionnaires d'information restent en charge de la production des métadonnées - en langage EML), des rôles sont précisés (les gestionnaires d'information contribuent au redéveloppement d'un standard pour le réseau LTER tandis que les développeurs travaillent au développement d'un standard pour les sciences de l'environnement), des pratiques locales sont maintenues (les gestionnaires d'information s'échangent des solutions *ad hoc* et des outils « maison »)... De l'autre, le standard EML transforme le monde. Des identités d'acteurs sont redéfinies (les gestionnaires d'information sont reconnus comme développeurs – et non pas seulement comme « implémenteurs »), de nouvelles structures organisationnelles sont consolidées (les gestionnaires d'information sont désormais représentés par le comité NISAC), de nouveaux modes de travail sont proposés (un espace de collaboration entre les sites et le réseau)...

Implémenter EML ne consiste pas seulement à améliorer (*to upgrade*) un système technique préexistant, il consiste aussi et surtout à redéfinir l'infrastructure sociotechnique qui soutient cet enchevêtrement de pratiques techniques, sociales et scientifiques. Or ces redéfinitions ont d'importantes conséquences sur le plan social et organisationnel. Parce que les dispositifs techniques sont intimement imbriqués dans les structures locales de travail, parce que le standard EML « fonctionne » dans une certaine

configuration (technique, sociale, organisationnelle), son éaction requiert des changements d'ordre infrastructurel⁴⁹.

Bibliographie

Akrich, M., 1992, « The de-scription of technical objects », in W.E. Bijker & J. Law (Eds.), *Shaping technology/Building Society. Studies in Sociotechnical Change*, Cambridge, MIT Press, pp. 259-263.

Arzberger, P.W., Schroeder, P., Beaulieu, A., Bowker, G.C., Casey, K., Laaksonen, L., et al., 2004, « Promoting Access to Public Research Data for Scientific, Economic, and Social Development », *Data Science Journal*, 3 (29), pp. 135-152.

Baker, K.S., & Karasti, H., 2005, *The Long Term Information Management Trajectory: Working to Support Data, Science and Technology*, San Diego, SIO Report.

Berg, M., 1996, « Practices of Reading and Writing: The Constitutive Role of the Patient Record in Medical Work », *Sociology of Health & Illness*, 18 (4), pp. 499-524.

Berg, M., 1997, *Rationalizing medical work : decision-support techniques and medical practices*, Cambridge, MIT Press.

Berg, M., & Bowker, G.C., 1997, « The Multiple Bodies of the Medical Record: Toward a Sociology of an Artifact », *The Sociological Quarterly*, 38 (3), pp. 513-537.

Bowker, G.C., 2001, « Biodiversity Datadiversity », *Social Studies of Science*, 30 (5), pp. 643-684.

Bowker, G.C., 2006, *Memory Practices in the Sciences*, Cambridge, MIT Press.

Bowker, G.C., & Star, S.L., 1999, *Sorting things out : classification and its consequences*, Cambridge, MIT Press.

Callon, M., 1986, « Éléments pour une sociologie de la traduction : la domestication des coquilles St-Jacques et des marins pêcheurs dans la baie de St. Brieuc », *L'Année Sociologique*, 36, pp. 169-208.

Callon, M., 2001, « Actor Network Theory », in N.J. Smelser & P.B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences*, Amsterdam, New York, Elsevier, pp. 62-66.

⁴⁹ Nous tenons à remercier nos collaborateurs Karen Baker (Scripps Institution of Oceanography) et David Ribes (University of California, San Diego), tous deux membres de l'équipe de recherche formée autour du *Comparative Interoperability Project* dont cette étude est tirée, pour leur précieuse contribution.

- Fountain, J.E., 2001, *Building the Virtual State: Information Technology and Institutional Change*, Washington, Brookings Institution Press.
- Franklin, J.F., Bledsoe, C.S., & Callahan, J.T., 1990, « Contributions of the Long-Term Ecological Research Program », *BioScience*, 40 (7), pp. 509-515.
- Fujimura, J.H., 1988, « The Molecular Biological Bandwagon in Cancer research: Where Social Worlds Meet », *Social Problems*, 35 (3), pp. 261-283.
- Fujimura, J.H., 1991, « On Methods, Ontologies, and Representation in the Sociology of Science: Where Do We Stand? », in D.R. Maines (Ed.), *Social Organization and Social Process*, New York, Aldine de Gruyter, pp. 207-248.
- Hobbie, J.E., Carpenter, S.R., Grimm, N.B., Gosz, J.R., & Seastedt, T.R., 2003, « The US Long Term Ecological Research Program », *BioScience*, 53 (1), pp. 21-32.
- Hutchins, E., 1995, *Cognition in the wild*, Cambridge, MIT Press.
- Jones, M.B., Berkley, C., Bojilova, J., & Schildhauer, M., 2001, « Managing Scientific Metadata », *IEEE Internet Computing*, 5 (5), pp. 59-68.
- Karasti, H., & Baker, K.S., 2004, « Infrastructuring for the Long-Term: Ecological Information Management », in *Proceedings of the Hawaii International Conference on System Sciences (HICSS'37)*, Hawaii, Computer Society Press (CD-ROM), 10 p.
- Karasti, H., Baker, K.S., & Bowker, G.C., 2002, « Ecological Storytelling and Collaborative Scientific Activities », *SIGGROUP Bulletin*, 23 (2), pp. 29-30.
- Latour, B., 1987, *Science in action : how to follow scientists and engineers through society*, Cambridge, Harvard University Press.
- Latour, B., 1993, *Aramis ou l'amour des techniques*, Paris, La Découverte.
- Lemke, J.L., 2000, « Across the Scales of Time: Artifacts, Activities, and Meanings in Ecosocial Systems », *Mind, Culture, and Activity*, 7 (4), pp. 273-290.
- Olson, G.M., Atkins, D., Clauer, R., Weymouth, T., Prakash, A., Finholt, T., et al., 2001, « Technology to support distributed team science: The first phase of the Upper Atmospheric Research Collaboratory (UARC) », in G.M. Olson, T. Malone & J. Smith (Eds.), *Coordination Theory and Collaboration Technology*, Hillsdale, Lawrence Erlbaum Associates, pp. 761-783.
- Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B., & Stafford, S.G., 1997, « Nongeospatial Metadata for the Ecological Sciences », *Ecological Applications*, 7 (1),

Florence Millerand 10/21/07 12:09 PM
Commentaire: Les actes ont été publiés sur CD-ROM, il n'y a donc pas de numéros de pages. J'ai indiqué en revanche le nombre de pages.

pp. 330-342.

Millerand, F., 1998, « Usages des NTIC, les approches de la diffusion, de l'innovation et de l'appropriation (1re partie) », *COMMposite*, 98 (1), http://commposite.org/v1/98.1/articles/ntic_1.htm

Florence Millerand 10/21/07 12:05 PM

Commentaire: Il s'agit d'une publication dans une revue électronique, il n'y a donc pas de pages. J'ai indiqué l'URL.

Millerand, F., 1999, « Usages des NTIC, les approches de la diffusion, de l'innovation et de l'appropriation (2e partie) », *COMMposite*, 99 (1), http://commposite.org/v1/99.1/articles/ntic_2.htm

Florence Millerand 10/21/07 12:05 PM

Commentaire: Idem

Millerand, F., Ribes, D., Baker, K.S., & Bowker, G.C., 2005, « Comparative Interoperability Project: Collaborative Science, Interoperability Strategies, and Distributing Cognition », *communication au Workshop "Artifacts & Collectives"*, Lyon, France.

Florence Millerand 10/21/07 12:07 PM

Commentaire: Il s'agit d'une communication, plus précisément d'une présentation par affiche (poster), le texte n'a pas été publié.

Ribes, D., Baker, K.S., Millerand, F., & Bowker, G.C., 2005, « Comparative Interoperability Project: Configurations of Community, Technology, and Organization », in *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries*, New York, ACM Press, pp. 65-66.

Shapin, S., 1989, « The Invisible Technician », *American Scientist*, 77, pp. 554-563.

Schumpeter, J., 1934, *The Theory of Economic Development*, Cambridge, Harvard University Press.

Star, S.L., 1991, « 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., & Bowker, G.C., 2001, « How to infrastructure », in L.A. Lievrouw & S. Livingstone (Eds.), *Handbook of New Media. Social Shaping and Consequences of ICTs*, London, Thousand Oaks, New Delhi, Sage Publications, pp. 151-162.

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, pp. 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), pp. 111-134.

Star, S.L., & Strauss, A., 1999, « Layers of Silence, Arenas of Voice: The Ecology of

Visible and Invisible Work », *JCSCW*, 8 (1/2), pp. 9-30.

Strauss, A., 1993, *Continual Permutations of Action*, New York, Aldine de Gruyter.

Timmermans, S., 1998, « Mutual Tuning of Multiple Trajectories », *Symbolic Interaction*, 21 (4), pp. 425-440.

Timmermans, S., & Berg, M., 1997, « Standardization in Action: Achieving Local Universality through Medical Protocols », *Social Studies of Science*, 27 (2), pp. 273-305.

Timmermans, S., & Berg, M., 2003, *The Gold Standard. The Challenges of Evidence-Based Medicine and Standardization in Health Care*, Philadelphia, Temple University Press.

Weick, K.E., 1979, *The Social Psychology of Organizing*, Addison-Wesley.