



HAL
open science

S'hybrider sous le 4e paradigme de la “science dirigée par les données massives” : l'ouverture des données favorise-t-elle l'interdisciplinarité et la transdisciplinarité ?

Hervé Squvidant, Florence Revelin

► To cite this version:

Hervé Squvidant, Florence Revelin. S'hybrider sous le 4e paradigme de la “science dirigée par les données massives” : l'ouverture des données favorise-t-elle l'interdisciplinarité et la transdisciplinarité ?. Communication, Technologies et Développement, 2021, 9, 10.4000/ctd.3791 . hal-03185906

HAL Id: hal-03185906

<https://hal.inrae.fr/hal-03185906v1>

Submitted on 30 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

S'hybrider sous le 4^e paradigme de la "science dirigée par les données massives" : l'ouverture des données favorise-t-elle l'interdisciplinarité et la transdisciplinarité ?

*Hybridizing under the 4th paradigm of "data-intensive science": does open-data
facilitate interdisciplinarity and transdisciplinarity?*

*Hibridar bajo el cuarto paradigma de "intensiva en datos": ¿los datos abiertos
promueven la interdisciplinariedad y la transdisciplinariedad?*

Hervé Squidant et Florence Revelin



Édition électronique

URL : <http://journals.openedition.org/ctd/3791>

ISSN : 2491-1437

Éditeur

Chaire Unesco Pratiques émergentes en technologies et communication pour le développement

Édition imprimée

ISBN : 2491-1437

Ce document vous est offert par INRAE Institut National de Recherche pour l'Agriculture,
l'Alimentation et l'Environnement



Référence électronique

Hervé Squidant et Florence Revelin, « S'hybrider sous le 4^e paradigme de la "science dirigée par les
données massives" : l'ouverture des données favorise-t-elle l'interdisciplinarité et la
transdisciplinarité ? », *Communication, technologies et développement* [En ligne], 9 | 2021, mis en ligne le
23 mars 2021, consulté le 30 mars 2021. URL : <http://journals.openedition.org/ctd/3791>

Ce document a été généré automatiquement le 30 mars 2021.

Communication, technologies et développement

S'hybrider sous le 4^e paradigme de la "science dirigée par les données massives" : l'ouverture des données favorise-t-elle l'interdisciplinarité et la transdisciplinarité ?

Hybridizing under the 4th paradigm of "data-intensive science": does open-data facilitate interdisciplinarity and transdisciplinarity?

Hibridar bajo el cuarto paradigma de "intensiva en datos": ¿los datos abiertos promueven la interdisciplinarietà y la transdisciplinarietà?

Hervé Squidant et Florence Revelin

Introduction

- 1 Alimenter des entrepôts de données et contribuer à rendre disponibles des données de recherche, implique, pour les communautés qui les produisent (principalement des chercheurs, ingénieurs, doctorants ou étudiants), d'intégrer une projection dans les réutilisations de ces matériaux dans des cadres, des formes et des temporalités qui leurs sont inconnus, et par des acteurs qui peuvent être des humains ou des machines. A minima, cette opération suppose qu'ils se projettent dans l'objectif de rendre la donnée "appropriable" par autrui : c'est la finalité des principes FAIR² (Wilkinson et al., 2016) qui gouvernent aujourd'hui l'opérationnalisation de l'ouverture des données. Nous disposons à ce jour d'un faible recul sur la manière dont les communautés productrices de données se saisissent de ces enjeux et se projettent dans ce grand inconnu du réusage. En particulier, lorsque ces communautés rassemblent différentes cultures épistémiques (Knorr-Cetina, 1999), impliquant une pluralité de rapports à la

fois à la donnée, et à l'instrumentation technique et numérique en jeu pour intégrer l'objectif d'ouverture des matériaux de recherche.

- 2 Nous proposons dans cet article de suivre le cheminement de l'équipe d'un programme de recherche interdisciplinaire qui s'est confrontée à ces enjeux en abordant **le partage au sein de l'équipe scientifique, et plus largement l'ouverture des matériaux de recherche comme un levier pour dépasser des frontières épistémiques, à la fois disciplinaires, et entre sciences et société**. L'article retrace ainsi l'expérimentation menée par les chercheurs et ingénieurs impliqués dans le programme *Paroles et chemins de l'agriculture littorale* (Parchemins) **en partant de l'instrumentation numérique déployée en lien avec l'objectif de partage**. Ce programme a rassemblé, de 2016 à 2020, 10 scientifiques d'horizons multiples (agronomie, anthropologie, sciences et techniques de l'information, et études sociales des sciences) et de statuts divers³, équipe ponctuellement renforcée par des stagiaires et des partenaires extérieurs. Son objectif était de documenter et analyser les relations entre agriculture et littoral, et les transformations de l'agriculture sur les littoraux bretons depuis 70 ans. Outre sa vocation interdisciplinaire, ce projet est caractérisé par son ouverture vers la société (transdisciplinarité) : ses membres reconnaissent aux savoirs d'expérience un vrai statut dans la conduite de la recherche. Cela implique, pour l'équipe porteuse, d'une part une ouverture à la diversité des formes de problématisation opérées par les acteurs sociaux et une capacité à les restituer et les articuler avec ses propres questions de recherche, et d'autre part, une mise à disposition des productions intermédiaires (notamment les matériaux de recherche) au cours du projet dans une logique de partage et de co-construction. Cet engagement particulier a fait la part belle aux activités et réflexions liées à l'ouverture des données, et à son instrumentation technique et numérique. Une telle posture repose sur la double hypothèse qu'une approche intégrée des matériaux de recherche hétérogènes du programme, aidée par des outils hybrides, favorise la mise en œuvre de l'interdisciplinarité; et que l'ouverture de ces matériaux contribue à symétriser les rapports sciences-sociétés. A l'issue de cette expérimentation, nous proposons de revenir sur ces hypothèses de départ pour nous demander dans quelle mesure, dans le cadre d'un projet interdisciplinaire, **l'objectif de mise en partage des matériaux de recherche à un niveau à la fois interne et ouvert, favorise l'hybridation des approches scientifiques et l'ouverture vers la société. Comment l'instrumentation technique et numérique autour de l'opérationnalisation du partage des matériaux de recherche met-elle en lumière ces formes d'hybridation?**

Matériels et Méthodes

- 3 L'article est à l'image de l'expérience qu'il décrit : il est fondé sur une approche hybride, conjuguant l'approche matérielle et technique des sciences et technologies de l'information (STI) et la démarche ethnographique, qui prend appui sur cette finesse descriptive par les outils pour aborder la diversité des points de vue et dynamiques de collaborations qui se rattachent au processus d'instrumentation. La description de ce processus repose sur son auto-ethnographie⁴. Elle s'appuie sur l'observation participante et la collecte des traces du travail individuel et collectif aux différentes étapes de la conception ou des choix d'outillage numérique au service du partage des matériaux de recherche. Un corpus d'entretiens réalisés auprès des principaux membres du programme directement impliqués dans ce processus (8 entretiens avec

les chercheurs et ingénieurs du programme et 2 entretiens avec des stagiaires en anthropologie) complète ces observations.

- 4 La première partie rappelle le contexte général dans lequel le projet s'est développé, à savoir celui de l'avènement de la donnée massive et ouverte, qui renouvelle les cadres du travail scientifique. Dans ce contexte, nous exposerons les intentions et postures qui ont présidé à l'intégration de l'objectif de partage et d'ouverture des matériaux de recherche par l'équipe Parchemins, et donc guidé ses choix d'instrumentation numérique. La seconde partie revient sur les négociations autour de la co-conception d'un écosystème d'outils, entre les concepteurs et les chercheurs considérés comme premiers utilisateurs des outils numériques et des données partagées. La troisième partie propose une description fine des choix techniques qui ont été réalisés, et des outils mobilisés. Nous reviendrons en discussion sur la manière dont le collectif s'est hybridé au sein de ce processus d'instrumentation.

Mener un projet de recherche transdisciplinaire à l'ère de l'avènement de la donnée massive et ouverte

- 5 L'expérience vécue par l'équipe du programme Parchemins s'inscrit dans le contexte actuel de l'avènement de la donnée massive et ouverte, qui touche l'ensemble des communautés de recherche, au point de transformer profondément les pratiques de recherche et les formes de collectivisation du travail scientifique (Jaton et Vinck, 2014; Millerand, 2011; Oberhauser, 2014). Certains auteurs confèrent à ce bouleversement la portée d'un changement de paradigme. En 2009, Tony Hey et ses collaborateurs (Hey et al, 2009) ont ainsi publié un ouvrage collectif reprenant le concept attribué à Jim Gray, informaticien et chercheur au sein de l'entreprise américaine Microsoft, du Quatrième paradigme de la science : celui d'une science dirigée par l'intensification des données ("*data-intensive science*")⁵. Cette proposition s'appuie sur une vision évolutive et cumulative de la démarche et des méthodes scientifiques au gré des progrès de leur équipement technique. Gray compare ainsi les effets de la révolution depuis les années 2000 à ceux provoqués par l'invention de l'imprimerie au 15^e siècle. Le "quatrième paradigme" repose ainsi sur l'idée que la massification extrême des données submerge les capacités d'analyse des chercheurs "humains" et favorise de nouvelles formes de production des connaissances basées sur des algorithmes et des intelligences artificielles, portées par des instrumentations numériques en capacité de traiter ces masses de données intensivement alimentées. Si la lecture proposée par Gray peut être débattue, elle interroge et recentre l'attention sur les relations entre les données et l'instrumentation numérique à l'appui de leur circulation et leur appréhension (massive) comme facteurs de transformation de la démarche scientifique.
- 6 Cette attention accrue à la circulation des données numériques converge avec le mouvement contemporain de la "Science Ouverte" (Open Science) qui prône la mise à disposition en accès libre des publications scientifiques (*open access*) et des données (*open data*) (Plan pour la Science ouverte, 2018). La politique internationale d'ouverture des données repose sur des grands principes de transparence, d'efficacité d'une recherche cumulative, de reproductibilité, et de démocratisation de l'accès aux savoirs (*ibid.*). Elle a été forgée dans le contexte de l'intensification de la production des données sur le modèle de disciplines où la production des données est quantitative (telles que l'astronomie, la physique, ou la génétique par exemple), et en partie automatisée (Broudoux et Chartron, 2015). Elle véhicule une injonction à publier sur

des entrepôts de données et en accès ouvert, c'est-à-dire sous licence ouverte telle que celle proposée par *Etalab* (Licence Ouverte V 2.0, Etalab 2017), tous les matériaux de recherche produits sur la base de financements publics et non soumis à des régimes de protection juridique. Elle contribue ainsi à alimenter massivement les flux du "Big data" : le principe étant que, sauf injonction contraire, aucune donnée n'est vouée à rester cloisonnée et protégée dans un outil de stockage "fermé". Elle doit être mise en circulation libre sur le web dans le respect des principes FAIR pour être rendue facilement et mondialement appropriable.

- 7 À travers l'expérience que nous relatons, notre attention se porte non pas sur la conception de formes automatisées de production de données massives, mais plutôt sur un contexte de production de données plus artisanal, où l'enjeu pour la communauté de recherche n'est pas d'appréhender l'abondance des données, mais d'intégrer l'ouverture de celles-ci dans l'activité de recherche. Ainsi, les motifs, les valeurs et le champ d'application reliés à cette intention ne sont pas nécessairement ceux portés en priorité par ce double mouvement de massification/intensification et d'ouverture, même si *in fine*, elle l'alimente. Dans le cas étudié, il s'agit avant tout d'une posture engagée du collectif, concevant les productions scientifiques comme un "bien commun" et portant un engagement en faveur du "libre" (*e.g.* libre accès aux données, aux publications, licences libres). Par ailleurs, l'équipe du programme compte des chercheurs en agronomie et des ingénieurs en géomatique et STI, aguerris à la mise en partage de données (spatiales et environnementales) rendue obligatoire depuis 2007 par la directive Inspire (INSPIRE, 2007). Forts de cette expérience, le partage des matériaux de recherche y est envisagé comme support de la démarche interdisciplinaire, par extension des pratiques de mise en bases de données aux productions des chercheurs en sciences humaines et sociales (SHS), et l'ouverture des matériaux de recherche comme un potentiel pour tisser des liens entre sciences et société. Le mouvement de l'*open data* fournit ainsi un cadre de référence pour opérationnaliser cette intention, mais également pour expérimenter, éprouver et interroger les problématiques et limites soulevées par cet effort de mise à disposition libre et massive des matériaux de recherche "ouvrables", en particulier ceux issus de disciplines où le partage est loin d'être une pratique courante, comme l'anthropologie sociale, qui s'appuie sur des données très qualitatives et confère à leur analyse une ambition plus interprétative qu'explicative (Olivier de Sardan, 2008).
- 8 Une des spécificités du programme Parchemins tient au fait qu'il mobilise des connaissances hétérogènes et des thématiques de recherche issues de champs disciplinaires différents. La première ambition du projet porte donc sur une mise en relation dialectique de ces différents champs afin de transcender les cadres disciplinaires classiques, conceptuels et méthodologiques (Galochet, 2008). Une autre ambition concerne le porter à connaissance et le partage de matériaux de recherche à la fois auprès d'autres chercheurs travaillant sur les relations entre agriculture et littoral, aujourd'hui ou dans le futur, mais également auprès d'acteurs hors recherche institutionnelle comme les organismes publics gestionnaires, mais aussi les agriculteurs et autres groupes socio-professionnels ou associatifs impliqués dans la compréhension, la problématisation, la critique, la gouvernance de l'agriculture sur le littoral.
- 9 Cette double ambition inter et transdisciplinaire se traduit en particulier par l'identification du partage des matériaux de recherche produits par le programme comme un enjeu explicite de celui-ci, et complémentaire à l'objectif de formalisation

des analyses et production de connaissances aux formats académiques classiques. Dans ce contexte, l'équipe Parchemins a fait le choix d'expérimenter la conceptualisation, l'élaboration, le déploiement et l'alimentation d'un système d'information (SI) dédié proposant la publication et la mise en partage des données et autres connaissances produites dans le cadre du projet. Au démarrage du projet, les projections sur l'outillage restent abstraites et renvoient avant tout à l'intention de mise en circulation des données, comme l'illustre le croquis réalisé par une ethnologue du programme en 2017 (figure 1), qui imagine quatre phases : la production, la publication, la réutilisation et le rapportage. D'emblée, ce SI n'a pas été imaginé ou perçu simplement comme un outil ou un ensemble d'outils au service de la recherche, mais comme un objet de recherche en tant que tel, soutenant l'effort de réflexivité et d'auto-enquête lié au processus d'instrumentation du partage des données : il s'agit, par exemple, de rendre possible la mobilisation par une agronome de "données" d'enquêtes ethnographiques qu'elle n'a pas produites; ou, à l'inverse, à un ethnologue d'utiliser des données statistiques géolocalisées, dans le cadre de leurs analyses respectives. En quoi chaque option et choix réalisés lors des différentes étapes de la conception de ce SI conditionnent-ils la construction d'une interdisciplinarité effective ou pas ? Comment les outils conditionnent-ils, fabriquent-ils, interrogent-ils, répondent-ils à cette recherche interdisciplinaire ?

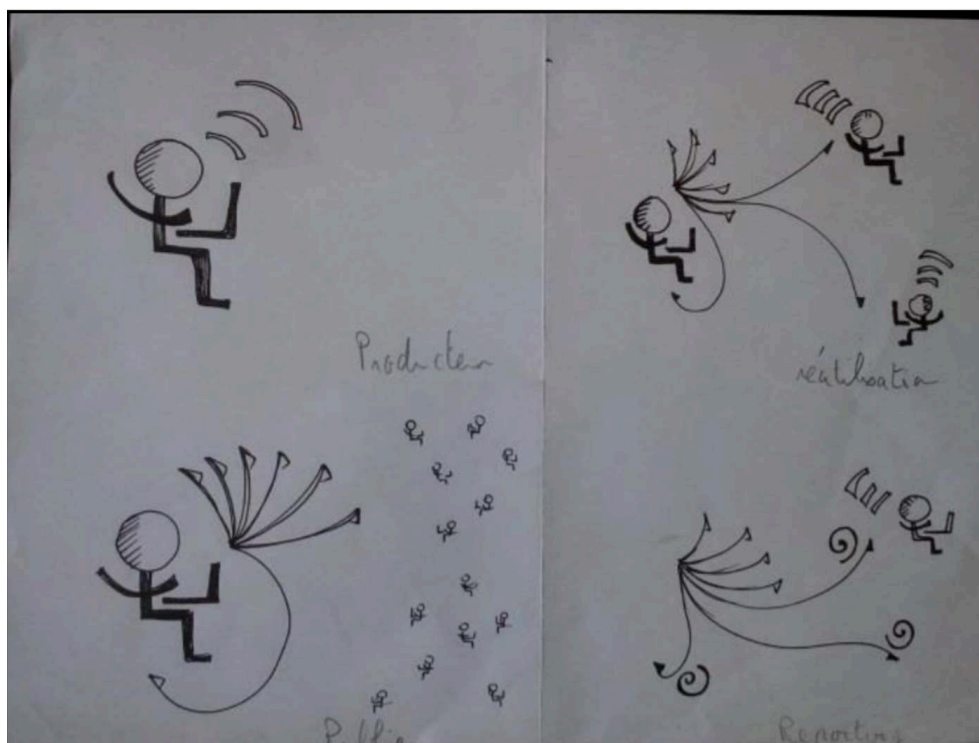


Figure . Croquis représentant la circulation des données telle qu'imaginée au début du programme, réalisé par Marine Legrand, anthropologue. 2017.

Négocier l'instrumentation numérique : d'un outillage statique à un écosystème dynamique

- 10 Instrumenter la mise en partage des données du programme n'a pas été simple et s'est heurté à différents écueils. Le premier d'entre eux est une différence de vision sur la

question du numérique au service des données entre les concepteurs du système d'information et les chercheurs. Les deux ingénieurs en STI relatent une expérience préalable similaire : quels que soient leurs champs disciplinaires, les chercheurs avec qui ils collaborent expriment spontanément leurs besoins d'instrumentation numérique par le fait de concevoir une base de données, conçue comme une fin en soi. Face à cette demande, l'ingénieur en STI a deux réactions possibles en fonction de sa posture vis à vis du projet: soit il reçoit cette injonction émanant généralement de la hiérarchie et l'applique comme telle, soit il peut initier un réel échange entre les deux parties dans l'idée de co-concevoir un outil qui embarque une forme d'innovation et d'adaptation aux besoins et spécificités du projet. Ce type d'initiative tend à sortir d'une conception fondée sur un outil unique, et souvent répliquable dans différents projets avec relativement peu de souplesse technique, pour aller vers une conception plus ouverte et souple : celle d'un système d'information dédié aux besoins du projet et composé d'un ensemble de services.

- 11 Le projet Parchemins n'a pas échappé à cette tendance générale: la demande initiale des chercheurs était celle d'une base de données, qui a évolué, sous l'impulsion des concepteurs, vers la co-conception de ce qu'ils nomment un "écosystème de services". Le projet d'instrumentation numérique est ainsi passé d'une vision que l'on peut qualifier de statique de la gestion de la donnée à une approche dynamique du "chemin de la donnée" intégrée à des services web accessibles de manière simple et interopérable pour les collecteurs et les utilisateurs de ces données, suivant ainsi le modèle "Data as a service" (Daas) (Truong et al., 2009).
- 12 L'hétérogénéité des connaissances produites, d'une part, et l'expérience acquise par les concepteurs dans la publication des données émanant des sciences de l'environnement d'autre part, ont fourni des arguments en faveur de ce choix. Ces arguments ont souligné la pertinence à concevoir un système d'information qui soit à la fois :
- 13 **hybride**, pour intégrer des matériaux de recherche de natures et formats très hétérogènes;
- 14 **modulaire**, pour offrir une souplesse dans l'idée de tester simplement de nouveaux outils indépendamment les uns des autres et de faire ainsi évoluer le SI;
- 15 **interopérable**, le respect des standards d'échanges des données permettre les interactions en interne entre les différents modules du SI et en externe au sein de l'écosystème numérique préexistant;
- 16 **libre**, pour garantir l'adaptabilité de ce type d'outil grâce à l'ouverture du code source. Ils permettent également de maîtriser complètement certains aspects importants comme le lieu de stockage des données.
- 17 Les principes actés par l'équipe (et en particulier la coordinatrice du programme et les ingénieurs concepteurs) étaient que ce SI s'appuierait sur des services préexistants au projet comme l'Infrastructure de Données Spatiales (IDS) GéoSAS (Béra, 2015), qui permet la diffusion et le partage des données géographiques conformément à la directive européenne INSPIRE (INSPIRE,2007) en respect des standards de l'interopérabilité. Il serait progressivement complété par un certain nombre de modules proposant par exemple la publication et le partage des données ethnographiques et aussi la possibilité de compiler au sein d'un même outil les matériaux de recherche émanant des différents champs disciplinaires. Le choix des

modules à concevoir, à développer ou simplement à réutiliser s'ils existaient déjà, devait dépendre :

- 18 Du respect des bonnes pratiques émanant de l'expérience acquise dans le monde des sciences de l'environnement : modularité, interopérabilité, logiciel libre, etc
- 19 De leur conformité aux nouveaux principes du FAIR,
- 20 De leur capacité à répondre aux multiples questions et débats au sein de l'équipe scientifique, notamment ceux qui concernent les dimensions éthiques et juridiques,
- 21 De leur faculté à favoriser l'interdisciplinarité en intégrant l'ensemble des matériaux de recherche produits,
- 22 De leur capacité à être facilement appropriés par les chercheurs et par un public plus large.
- 23 Pour avancer dans la concrétisation du SI, les ingénieurs en STI ont réalisé une veille et une étude prospective sur les outils existants. Ils ont distingué trois grandes voies de développement et d'intégration des outils dans le projet de SI :
- 24 des outils et des services web préexistants émanant de différentes communautés (recherche, GAFA, ..) déjà déployés et prêts à l'emploi, mais qui peuvent poser le problème d'être propriétaires,
- 25 des logiciels libres développés selon les principes de l'Open Source, que l'on peut télécharger et déployer dans l'infrastructure,
- 26 des outils "maison" à concevoir, à développer et à déployer.
- 27 L'équipe d'ingénieurs a ainsi tenté d'identifier des outils ou des services numériques répondant aux intentions initiales du projet, dans l'idée de les tester au sein du SI Parchemins.

Collaborations, arbitrages et itération autour des outils de partage des données

- 28 Une fois ces grands principes de mise en place du système d'information établis et validés collectivement, est venu le moment des arbitrages et des choix pour les différents outils numériques. Cette phase s'est traduite par un processus itératif de questionnements d'ordres techniques et informatiques sur les fonctionnalités d'une part, portés par les concepteurs du SI, c'est-à-dire deux ingénieurs de recherche en sciences et technologies de l'information, et une ingénieure de recherche en anthropologie et d'autre part, les attentes *a priori* non définies et peu formalisées des futurs utilisateurs, à savoir l'ensemble des chercheurs impliqués dans le programme.
- 29 Chaque outil a ainsi été questionné au regard des fonctionnalités supposément attendues par les utilisateurs d'une part et du respect des grands principes prédéfinis qui imposent le fait qu'ils soient tous libres, modulaires et interopérables d'autre part. Cette phase d'aller-retour entre concepteurs et utilisateurs de première ligne a été complexe et poreuse : les besoins des utilisateurs se sont avérés difficiles à formuler précisément, rebondissant continuellement sur les possibilités concrètes offertes par des outils qui leur apparaissent comme abstraits. L'expérience et la veille réalisée par un ingénieur en STI a eu un caractère déterminant, en ce qu'elles ont permis de proposer des solutions et des exemples concrets à expérimenter pour donner prise aux projections des chercheurs dans l'usage des outils. Cette concrétisation par l'exemple

s'est avérée une étape absolument nécessaire du point de vue des utilisateurs de première ligne.

- 30 Le premier module expérimenté a été un service de stockage et de partage de fichiers, initialement réservé aux membres du projet. Avant le lancement du projet de conception de SI, l'équipe utilisait l'outil propriétaire Dropbox pour partager des documents de travail. Ce choix avait été fait par défaut, en continuité avec des pratiques déjà instaurées pour plusieurs membres du programme, qui ont spontanément élargi l'usage de cet outil déjà connu et utilisé au collectif. Cette solution est simple à mobiliser puisqu'il suffit de créer un ou plusieurs comptes et de partager les dossiers et les fichiers. Elle est cependant limitée et pose des questions sur le plan éthique, juridique et pratique. Le volume maximum utilisé est de 2 Giga-octets et les données sont stockées aux USA par la société DropBox échappant donc à tout contrôle, et en particulier au champ d'application du RGPD⁶. C'est sans doute l'argument éthique qui a emporté le choix de l'équipe d'évoluer vers la solution Owncloud, déployée dans l'infrastructure via des serveurs hébergés au sein du service informatique d'une des tutelles de l'équipe. La limite de stockage correspondait au volume de stockage du serveur, c'est-à-dire plusieurs Tera-octets. Le "cloud", affectueusement surnommé "Jean-Cloude", a été énormément utilisé: en interne d'abord, puis aussi en externe pour partager des documents dans le cadre d'actions comme la production d'émissions radiophoniques ou l'organisation des Rencontres de Kervic (rencontres sciences-société sur le thème de l'agriculture littorale), en ouvrant un compte aux partenaires de Parchemins (appelés les "compagnons"), pour leur permettre d'accéder à certains dossiers en rapport avec ces activités. L'équipe a également exploré l'usage du cloud comme outil de publication de certaines données en utilisant les fonctions de partage d'Owncloud.
- 31 La volonté de partager en mode ouvert les matériaux ethnographiques collectés ainsi que les données quantitatives au travers d'un service web unique interopérable et ergonomique a orienté l'équipe vers la solution CKAN⁷ (Pollock et al., 2009). Cet outil est la référence des portails "Open data" officiels dans la majorité des pays occidentaux ; c'est un catalogue générique, qui n'est pas dédié spécifiquement aux données de recherche. Il répondait donc *a priori* à l'ambition transdisciplinaire, puisqu'il était capable de diffuser n'importe quel type de données numériques sous la forme de fichiers vidéo, son, image, tableur, carte, texte, rapport, etc... Son interface est aussi adaptée à des usages experts et permet l'indexation et la recherche des données publiées selon des thématiques scientifiques. Ces fonctionnalités ont cependant nécessité des développements spécifiques pour intégrer le thésaurus thématique du projet : un système d'indexation co-construit par l'ensemble de l'équipe pour couvrir l'inégalité des matériaux de recherche et les indexer leur contenu sur la base des thématiques de recherche du programme. Le processus de constitution de ce thésaurus a été long et fastidieux, du fait de la complexité de la mise en relation dialectique des différents champs disciplinaires représentés au sein du projet. CKAN permet en principe, en parallèle, un usage "grand public" grâce à la simplicité des recherches de type "Google" (*i.e. via* un moteur) complétées par le filtrage des réponses selon le type ou le format des données, le site d'étude, les mots-clés, le type de licence, etc... Enfin, son interopérabilité simplifie son intégration au sein de l'écosystème externe par le mécanisme de moissonnage des métadonnées, à la fois en les exportant vers des

catalogues tiers et en important des métadonnées de type INSPIRE à partir du catalogue de GéoSAS.

- 32 Dès le départ du projet a émergé le besoin d'un outil communautaire pour porter à connaissance en interne et en externe les actions et les productions réalisées au cours du programme, qui, comme de nombreux projets de recherche, ne disposait que de peu de moyens pour recourir à des prestations de services externes. Les fonctions de blog et de communication éditoriale ont orienté l'équipe vers la solution WordPress, qui permettait de déployer un site web ou CMS (Content Manager System) dédié au projet, ainsi que des interactions avec ses visiteurs par le biais de la fonctionnalité "commentaires". Le site⁸ a ensuite évolué de telle sorte qu'il est devenu la porte d'entrée du Système d'information et propose des liens vers les différents modules qui le composent (notamment un cloud, un catalogue CKAN, un catalogue GéoSAS et un outil de visualisation d'indicateurs agronomiques spatialisés Mviewer). Dans sa structure initiale, il est le fruit d'une co-construction émanant de l'ingénieure en anthropologie et d'une ingénieure en base de données, partenaire du projet la première année. Il a ensuite été alimenté par tous les membres du projet, principalement sous la forme d'articles éditoriaux.
- 33 Finalement, en optant pour un écosystème d'outils modulaires, l'équipe envisageait de se prémunir face au piège du déploiement d'une instrumentation numérique dans la temporalité d'un programme. Un risque souvent encouru est en effet que les outils ne deviennent opérationnels que dans les phases finales du programme, car le processus d'instrumentation adaptée nécessite du temps. Néanmoins, si l'on schématise, il aura fallu environ la moitié de la durée du programme pour stabiliser l'ensemble des choix techniques, la seconde moitié étant dédiée au déploiement, à la mise en service et à la prise en main. Cette expérience témoigne du caractère profondément itératif et dynamique de la co-conception d'un SI dédié à une recherche scientifique, en particulier dans des contextes où des matériaux de recherche et des cultures épistémiques très hétérogènes entrent en confrontation dans ce processus d'instrumentation. Un second croquis, réalisé par une ingénieure du programme en 2019 à des fins pédagogiques, illustre la version stabilisée du SI telle qu'elle a été déployée et utilisée sur la fin du programme.

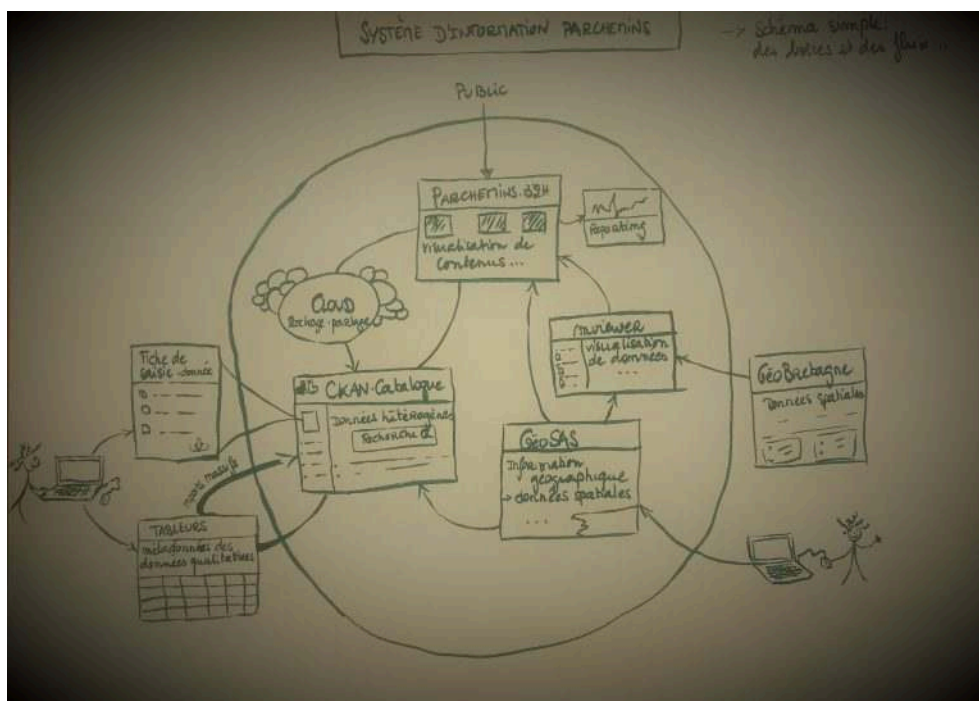


Figure . Croquis représentant les outils modulaires composant le système d'information du programme Parchemins tel que déployé en avril 2019, réalisé par Florence Revelin, anthropologue.

Discussion : quels dépassements des frontières épistémiques ?

- 34 Les choix qui ont présidé à la mise en place du système d'information s'inscrivent dans un processus fait de négociations et de tâtonnements. Cette recherche d'une instrumentation numérique adaptée aux besoins à la fois individuels (propres à chaque chercheur, en lien avec son origine disciplinaire et ses pratiques individuelles) et collectifs ne peut être dissociée d'un dialogue entre les chercheurs, dont les pratiques, les méthodes, et les matériaux de recherche sont hétérogènes et reposent sur des épistémologies différentes. Définir et co-construire, entre chercheurs d'horizons différents et ingénieurs en STI, les outils numériques adaptés pour prendre en compte cette hétérogénéité des cultures épistémiques (Knorr-Cetina, 1999) en présence représente un défi. Cet objectif commun - s'équiper pour partager les matériaux de recherche hétérogènes - constitue un cadre privilégié de confrontation entre ces cultures épistémiques variées (Millerand, 2011).
- 35 On peut ainsi décomposer ce *travail collectif scientifique et technique* - c'est-à-dire l'objet des interactions de l'équipe autour de la conception du SI Parchemins - en trois grands registres complémentaires. D'abord un **registre technique et ergonomique**, en référence à la définition des outils et modules décrits en première partie : il s'agit de définir des fonctionnalités, de sécuriser le stockage, d'optimiser l'ergonomie des outils, de favoriser le principe de subsidiarité⁹. Ensuite, un **travail sur les données collectées** : sélection, traitement, formatage. Et enfin, un **travail de documentation des données vouées à être mises en partage** au sein de certains modules du SI, notamment le catalogue. Il s'agit de définir des métadonnées qui décrivent les données publiées, à la fois sur la forme (format, contexte et date de collecte par exemple) et sur le fond (titre, résumé éventuel, mots-clés), et qui respectent à la fois les standards

d'interopérabilité et les exigences scientifiques et éthiques des chercheurs. Ce travail de documentation implique une caractérisation fine du statut juridique et éthique de chaque donnée, qui va conditionner son degré d'ouverture possible. Ces différents registres sont empreints d'une dimension politique, en référence aux engagements et choix qui sont constamment négociés puis arbitrés à chacune des étapes que nous avons décomposées. Au long des 4 années du programme *Parchemins*, l'équipe scientifique et technique s'est mise en mouvement à ce triple niveau. Ce processus a confronté les différentes cultures épistémiques en présence, dans le cadre de ce que Florian Jaton et Dominique Vinck (2016) décrivent comme des "processus frictionnels". Ils constituent des cadres dans lesquels les chercheurs et ingénieurs sont questionnés de multiples manières sur leurs pratiques et leurs rapports aux matériaux de recherche qu'ils produisent et gèrent, au prisme des enjeux et problèmes soulevés par les différents registres de travail collaboratif autour du SI. Ces processus génèrent des collaborations nouvelles (Granjou et al. 2014) et suscitent de nouvelles formes de collectivisation du travail scientifique (Browker et al. 2010; Jaton et Vinck, 2016).

Construire un langage commun autour du partage des données

- 36 Confronter les cultures épistémiques variées de l'équipe implique en premier lieu de rendre possible une intercompréhension entre les domaines de recherche et spécialités de chacun. Rattachés à des horizons disciplinaires divers et produisant des matériaux de recherche tout aussi divers, les membres du programme ne parlaient pas le même langage lorsqu'ils évoquaient les "données" de recherche, une problématique qui devient fondamentale lorsqu'il s'agit d'envisager leur partage effectif. Ces écarts de langage traduisent des décalages plus profonds qui renvoient aux catégories *emic*¹⁰ des différents membres du programme, reposant sur des conceptions et des épistémologies différentes. C'est donc le cas de manière emblématique pour le terme de « donnée ». Les contributions de l'histoire et de la sociologie des sciences aux critiques de la conception objectivante que sous-tend la notion de "donnée", mettent en évidence son impossible dissociation du processus par lequel des informations sont construites (Gitelman et al., 2013). Des constructions qui impliquent nécessairement le geste du chercheur et ne peuvent être conçues comme *données*, mais plutôt comme "obtenues" (Latour, 2007) ou "*capta*" en anglais (Drucker, 2011). Alors que cette notion est communément utilisée par les chercheurs en sciences agronomiques et les ingénieurs en STI, et qu'il s'agit du terme consacré dans le mouvement de la Science Ouverte, son usage est beaucoup plus parcimonieux dans le contexte des SHS et de l'anthropologie. Si les ethnologues du programme tendaient initialement à évoquer leurs "matériaux de terrain", ou plus spécifiquement leurs "corpus d'entretiens", leurs "observations", leurs "photos", ou leurs corpus de "délibérations" par exemple, le travail collaboratif en contexte interdisciplinaire a réifié le terme de "données" par effet d'emprunt et d'usage commun. Il n'en reste pas moins que lorsqu'il s'est agi de circonscrire ce qui, dans leurs corpus, fait "donnée" - au sens de ce qui est mis en partage dans le SI et documenté par des métadonnées - la tâche s'est avérée complexe et fastidieuse, faisant resurgir ces écarts épistémologiques.
- 37 Un autre niveau d'hybridation est celui de l'inclusion au sein du collectif, par le biais des ingénieurs en STI, du langage de l'informatique et du numérique. En se situant dans une démarche de conception et déploiement d'outils de partage des données, le collectif est entré en contact avec des enjeux techniques et numériques soulevés par les

outils. Si le travail et les compétences des chercheurs ne se substituent pas à ceux des ingénieurs qui restent pilotes sur ce registre, un niveau d'acculturation demeure nécessaire pour pouvoir rendre les arbitrages qui se présentent à divers moments du processus. Ils concernent par exemple le choix des outils et des fonctionnalités que nous avons évoqués, les standards de métadonnées qui seront retenus pour permettre une interopérabilité (par exemple DDI ou Dublin Core), ou les méthodes et procédures d'alimentation du catalogue ou des différents modules (par import massif ou saisie directe). D'autre part, ce niveau d'acculturation peut être lié à la prise en main de certains outils. C'est le cas pour les langages informatiques utilisés pour le site du programme (PHP sous WordPress) et pour le catalogue (Markdown pour Ckan) qui sont directement alimentés par les chercheurs. Sans couvrir exhaustivement la portée et les limites de cette acculturation au langage, et plus largement au monde de l'ingénieur en STI, l'expérience Parchemins montre qu'une appropriation basique et au moins superficielle liée au fonctionnement des outils est nécessaire à l'échelle du collectif. Elle peut représenter un coût important pour les chercheurs peu familiers de cet univers. L'ajustement inverse est vrai aussi : pour proposer et co-concevoir l'instrumentation numérique, l'ingénieur STI doit se plonger et se projeter dans le travail des chercheurs et saisir leurs rapports différenciés aux matériaux spécifiques qu'ils produisent, dans leurs dimensions pratiques, épistémologiques et éthiques.

Des cadres d'explicitations multiples pour réduire les écarts épistémiques

- 38 Outre la construction d'un langage commun, un temps de socialisation, de traduction et d'appropriation a été nécessaire pour rapprocher les différentes cultures épistémiques en jeu. Cela s'est produit de plusieurs manières : d'abord donc, par l'identification d'écarts plus ou moins manifestes que le langage permet facilement de saisir, ensuite par l'échange de bibliographie sur des notions ou des concepts clés, par l'élaboration de documents de travail internes ou des notes publiées sur le site (par exemple une brève expliquant ce qu'est une « donnée ethnographique »¹¹ qui a d'abord circulé comme document de travail interne), et enfin, majoritairement par des temps d'échange programmés ou fortuits, nécessaires à des formes d'explicitation. La mise en place du SI du programme a ainsi constitué un catalyseur de cet effort d'explicitation, et concentré une attention accrue au processus de fabrication de la recherche inter- et transdisciplinaire. Ce processus nécessite en effet d'explicitier très précisément comment et sur quoi chaque chercheur travaille, ce qu'il produit comme matériaux de recherches ou « données », comment et dans quelles mesures ces matériaux peuvent être appropriés ou utilisés par d'autres. Sans retracer en détail l'étendue et l'intensité des ajustements qui se sont produits, soulignons que cette co-construction fut un travail de longue haleine, souvent exprimé dans les témoignages des membres de l'équipe sur le registre de la "souffrance", en lien avec la dimension itérative du processus qui nécessite une endurance sur le long terme, et le coût parfois élevé des ajustements réalisés par chacune.e.

Conclusion

- 39 L'instrumentation numérique ne favorise donc pas à proprement parler l'interdisciplinarité, mais crée un cadre pour des formes d'hybridation des cultures épistémiques. Le chemin nécessairement parcouru dans la co-conception d'un écosystème de services numériques compatible avec le mouvement de la science ouverte contribue à construire un langage commun autour des données. Il oblige à se familiariser avec des rapports hétérogènes à des matériaux de recherche (qui ne pourraient être froidement versés dans des outils "clé en main" mais nécessitent au contraire d'aborder la complexité des épistémologies qui ont présidés à la production), des formes que ces matériaux ont ou peuvent prendre, et des rapports complexes qui les relient à leurs producteurs. En cela, l'expérience collective de l'équipe Parchemins ne produit pas une interdisciplinarité dans le sens initialement envisagé, c'est-à-dire à partir du partage des données, mais par cet itinéraire de co-construction qui fait émerger une question de recherche commune autour de la donnée et de son instrumentation. Cet itinéraire résulte nécessairement d'une collaboration dans la durée, effective et itérative entre les chercheurs et les ingénieurs impliqués dans cette volonté d'ouverture des matériaux de recherche. En cela, il reconfigure les rapports entre les métiers de la recherche dans la dialectique scientifique et technique qu'imposent ces processus et interpelle sur la diversité des statuts dans un contexte de contractualisation croissante. Par ailleurs, si cette ouverture des matériaux de recherche contribue, de fait, au mouvement de massification du « big data » tel que décrit par Jim Gray, elle contribue aussi à tisser des liens entre sciences et société dans une approche plus artisanale au service de la transdisciplinarité, en arpentant ce que les ingénieurs du programme conçoivent comme "les chemins et les circuits courts de la donnée", sous réserve qu'elle s'appuie sur des outils permettant une double entrée amatrice et experte. Il s'agit là d'une démarche nécessairement engagée, qui repose sur la manière dont l'équipe scientifique et technique se saisit de l'objectif d'ouverture et par laquelle elle lui donne un sens en pratique.

BIBLIOGRAPHIE

- Bera, R., Squidant, H., Le Henaff, G., Pichelin, P., Ruiz, L., Launay, J., Vanhouteghem, J., Arousseau, P., et Cudennec, C. (2015). : GéoSAS : A modular and interoperable Open Source Spatial Data Infrastructure for research, *Proc. IAHS*, 368, 9–14.
- Broudoux E., Ghislaine C. (2015). *Big Data - Open Data : Quelles valeurs ? Quels enjeux ?* Paris : De Boeck Supérieur.
- Bowker, G.C., Baker, K., Millerand, F. et Ribes, D. (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment, in : Hunsinger J., Klastrup L. et Allen M. (dir.), *International Handbook of Internet Research* (p. 97-117). Dordrecht : Springer.
- Drucker, J. (2011), Humanities Approaches to Graphical Display, *Digital Humanities Quarterly*, 5(1).

- Galochet M., Longuépée J., Morel V. et Petit O. (2008). *L'environnement : discours et pratiques interdisciplinaires*. Villeneuve d'Ascq : Artois Presses Université.
- Gitelman L. (2013). *Raw data is an oxymoron*. Cambridge : MIT Press.
- Granjou, C., Mauz, I., Barbier, M. et Breucker, P. (2014). Making taxonomy environmentally relevant. Insights from an all taxa biodiversity inventory. *Environmental Science & Policy*, 38,254-262.
- Hey T., Tansley S. et Tolle, K.. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Redmond (USA) : Microsoft Research.
- INSPIRE. (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14/03/2007 establishing an Infrastructure for Spatial Information in the European Community.
- Jaton, F. et Vinck., D. (2016). Processus frictionnels de mises en bases de données. *Revue d'anthropologie des connaissances*, 10 (4), 489-504.
- Knorr-Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. London : Harvard University Press.
- Latour, B. (1993). Le topofil de Boa-Vista. *Raisons pratiques*, 4, 187-216.
- Millerand, F. (2011). Le partage des données scientifiques à l'ère de l'e-science : l'instrumentation des pratiques au sein d'un collectif multidisciplinaire. *Terrains travaux*, 18 (1): 215-37.
- Oberhauser, P.-N. (2016). Des collaborations équivoques. *Revue d'anthropologie des connaissances*, 10 (4), 557-86.
- Olivier de Sardan, J.P. (2008). *La rigueur du qualitatif. Les contraintes empiriques de l'interprétation socio-anthropologique*. Louvain-La-Neuve : Academia
- Pollock, R. et Dietrich, D. (2009, décembre). CKAN : apt-get for the debian of Data. Communication présentée au 26^e Chaos Communication Congress. Berlin.
- Truong, H.-L. et Dustdar, S. (2009). On Analyzing and Specifying Concerns for Data As a Service, *Proceeding of IEEE Asia Pacific Services Computing Conference*, 87-94.
- Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 3.

NOTES

2. FAIR : “ Findable, Accessible, Interoperable, Reusable”, c'est-à-dire en français : Facile à trouver, Accessible, Interopérable et Réutilisable.
3. Deux chercheuses en sciences agronomiques, trois chercheuses anthropologues, deux chercheurs en études sociales des sciences (STS), ainsi que deux ingénieurs de recherche en géomatique et sciences et technologies de l'information (STI) et une ingénieure de recherche en anthropologie. La moitié de l'équipe a un poste titulaire au moment du programme, l'autre moitié est contractualisée sur plusieurs périodes financées par le programme.
4. Les auteurs de l'article sont des membres du programme.
5. Il est présenté comme succédant au paradigme de la “science computationnelle”, qui correspond aux évolutions technologiques des dernières décennies ayant permis aux machines de faire des simulations sur la base de modèles complexes, lui-même précédé par le paradigme de la “science théorique” qui a consisté à généraliser les observations empiriques (comme la loi de

Newton par exemple), et par le paradigme de la "science empirique" basée sur l'observation (Hey et al., 2009).

6. Règlement général sur la protection des données, 2018.

7. CKAN : Comprehensive Kerbal Archive Network. Adresse du catalogue déployé : www.data.agriculturelittorale.fr

8. Le site est accessible via deux adresses : www.parchemins.bzh et www.agriculturelittorale.fr

9. C'est-à-dire publier les données au plus près de leurs producteurs.

10. Ce concept anthropologique fait référence aux catégories endogènes des acteurs, c'est-à-dire la manière dont ils conçoivent et expriment leurs points de vue.

11. voir la brève : <http://www.parchemins.bzh/index.php/actu-1/> (consulté le 10/10/2020)

RÉSUMÉS

Nous avons encore peu de recul sur la manière dont les communautés productrices de matériaux de recherche se saisissent des enjeux de l'ouverture des données dans le contexte actuel du *Big Data* et de la Science Ouverte, dépeint par Jim Gray comme l'ère du nouveau paradigme de la "Science dirigée par les données massives" (Hey, 2009). En particulier, lorsque ces communautés rassemblent différentes cultures épistémiques, impliquant une pluralité de rapports à la donnée. En prenant appui sur l'expérience d'un programme de recherche sur les relations en agriculture et littoral en Bretagne (Parchemins, 2016-2020), nous décrivons comment une équipe de 10 chercheurs et ingénieurs issus d'horizons divers (agronomie, sciences et techniques de l'information, anthropologie et études sociales des sciences) a intégré cet objectif en lien avec son ambition interdisciplinaire et transdisciplinaire, et a co-construit un système d'information modulaire et interopérable pour y répondre. Dans quelles mesures l'objectif de mise en partage des matériaux de recherche à un niveau à la fois interne et ouvert, favorise-t-il l'hybridation des approches scientifiques et l'ouverture de la science vers la société? La description fine de l'instrumentation technique et numérique du partage des matériaux de recherche nous permet de mettre en lumière et de discuter une partie de ces formes d'hybridation.

We still have little knowledge on how communities producing research materials grasp the challenges of opening data in the current context of *Big Data* and *Open Science*, pictured by Jim Gray as the era of a new Scientific Paradigm : the "Data-intensive Science" (Hey, 2009). In particular, when these communities gather different epistemic cultures, involving a plurality of relationships to data. Based on the experience of a research program on agricultural and coastal relations in Brittany (Parchemins, 2016-2020), we describe how a team of 10 researchers and engineers from various backgrounds (agronomy, information and technology sciences, anthropology and science and technology studies) has integrated this objective in line with its interdisciplinary and transdisciplinary ambition, and has co-constructed a modular and interoperable information system in this regard. To what extent does the objective of sharing research materials at both an internal and open level promote the hybridization of scientific approaches and the opening of science to society? The fine description of the technical and digital instrumentation of sharing research materials allows us to highlight and discuss some of these forms of hybridization.

Todavía tenemos poca perspectiva sobre la forma en que las comunidades que producen materiales de investigación comprenden los desafíos de los datos abiertos en el contexto actual de *Big Data* y *Open Science*, retratado por Jim Gray como la era de lo nuevo. paradigma de la "ciencia intensiva en datos" (Hey, 2009). En particular, cuando estas comunidades reúnen diferentes culturas epistémicas, implica una pluralidad de relaciones con los datos. Basándonos en la experiencia de un programa de investigación sobre relaciones agrícolas y costeras en Bretaña (Parchemins, 2016-2020), describimos cómo un equipo de 10 investigadores e ingenieros de diversos orígenes (agronomía, ciencia y tecnología de la información, antropología y estudio social de la ciencia) ha integrado este objetivo en línea con su ambición interdisciplinar y transdisciplinar, y ha co-construido un sistema de información modular e interoperable para cumplirlo. ¿En qué medida el objetivo de compartir materiales de investigación tanto a nivel interno como abierto promueve la hibridación de enfoques científicos y la apertura de la ciencia a la sociedad? La descripción detallada de la instrumentación técnica y digital del intercambio de materiales de investigación nos permite resaltar y discutir algunas de estas formas de hibridación.

INDEX

Mots-clés : Partage des matériaux de recherche, ouverture des données, instrumentation numérique et technique, interdisciplinarité, transdisciplinarité

Palabras claves : Intercambio de materiales de investigación, apertura de datos, instrumentación digital y técnica, interdisciplinariedad, transdisciplinariedad

Keywords : Sharing of research materials, open data, digital and technical instrumentation, interdisciplinarity, transdisciplinarity

AUTEURS

HERVÉ SQUIDANT

UMR SAS 1069, Agrocampus-Ouest

FLORENCE REVELIN

UMR LISIS 9003, INRAE

Les auteurs remercient l'ensemble des membres du programme de recherche *Parchemins - Paroles et chemins de l'agriculture littorale (2016-2020)* qui ont contribué à construire ce retour d'expérience, et en particulier, Alix Levain, la coordinatrice, pour ses relectures éclairantes, et Anne-Gaëlle Beurier, pour son implication dans l'enquête et les réflexions à l'origine de cet article.