



HAL
open science

Données à penser. Enjeux pratiques et éthiques autour des données dans le montage de projets de recherche européens

Delphine Cavallo, Camille Noûs

► To cite this version:

Delphine Cavallo, Camille Noûs. Données à penser. Enjeux pratiques et éthiques autour des données dans le montage de projets de recherche européens. *Tracés: Revue de Sciences Humaines*, 2020, Les sciences humaines et sociales au travail (II). Que faire des données de la recherche?, 2019 (#19 hors-série), pp.101-113. 10.4000/traces.10793 . hal-02902993

HAL Id: hal-02902993

<https://amu.hal.science/hal-02902993>

Submitted on 27 Jan 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



Tracés. Revue de Sciences humaines

#19 | 2019

Les sciences humaines et sociales au travail (ii): Que faire des données de la recherche ?

Données à penser. Enjeux pratiques et éthiques autour des données dans le montage de projets de recherche européens

Thinking data. Practical and ethical issues for data in the setting up of European research projects

Delphine Cavallo et Camille Noûs



Édition électronique

URL : <http://journals.openedition.org/traces/10793>

DOI : 10.4000/traces.10793

ISSN : 1963-1812

Éditeur

ENS Éditions

Édition imprimée

Date de publication : 31 décembre 2019

Pagination : 101-113

ISBN : 979-10-362-0227-8

ISSN : 1763-0061

Ce document vous est offert par Aix-Marseille Université (AMU)



Référence électronique

Delphine Cavallo et Camille Noûs, « Données à penser. Enjeux pratiques et éthiques autour des données dans le montage de projets de recherche européens », *Tracés. Revue de Sciences humaines* [En ligne], #19 | 2019, mis en ligne le 22 juillet 2020, consulté le 27 janvier 2021. URL : <http://journals.openedition.org/traces/10793> ; DOI : <https://doi.org/10.4000/traces.10793>



Tracés est mis à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Données à penser. Enjeux pratiques et éthiques autour des données dans le montage de projets de recherche européens

DELPHINE CAVALLO
CAMILLE NOÛS

Cet article prend la forme d'un retour d'expérience sur le montage de deux projets de recherche européens et vise à décrire les enchaînements entre objectifs, contraintes et prises de décisions autour de la gestion des données¹. En prenant en compte ses corollaires techniques, budgétaires, éthiques et professionnels, il s'agit d'aborder des problématiques liées à la politique de recherche sur projets au niveau national et international.

Dans le cadre de mes fonctions de responsable de la diffusion numérique à TELEMMe², l'un des onze laboratoires de la Maison méditerranéenne des sciences de l'homme (MMSH, Aix-en-Provence), j'ai été sollicitée par les équipes en charge des projets européens à la MMSH, notamment « l'ingénierie Europe »³, afin de participer au montage de deux projets H2020⁴. Dans les deux cas, j'ai été mobilisée pour la coconception et la rédaction du volet communication, dissémination et exploitation des

1 Je remercie Juliette Galonnier, Stefan Le Courant, Anthony Pecqueux et Christelle Rabier de m'avoir invitée à participer à ce numéro.

2 Temps, espaces, langages. Europe méridionale, Méditerranée, UMR 7303 (université d'Aix-Marseille, CNRS).

3 Il s'agit de la personne en charge d'accompagner les chercheurs et chercheuses et équipes de recherche dans le montage et la mise en œuvre de projets sur financements européens. Son ingénierie réside dans sa maîtrise des appels, modes de financement, règles budgétaires, techniques d'écriture, etc., mais surtout dans sa capacité à faire converger les projets de recherche avec les attendus – souvent implicites – de la Commission européenne. L'intitulé « ingénierie Europe », terminologie utilisée par le CNRS, repose sur un raccourci volontaire, qui la rend plus directement identifiable parmi les différents personnels d'appui à la recherche.

4 Projets de recherche montés en réponse aux appels du programme européen pour la recherche et l'innovation, Horizon 2020 (2014-2020).

résultats de la recherche⁵, puis par glissement sur la gestion des données : le *data management plan* (DMP)⁶ et la partie éthique.

Ces deux projets étaient l'occasion pour moi de me confronter pour la première fois à la gestion des données de la recherche dans un cadre européen. Avant de rejoindre TELEMM, j'étais responsable du pôle Information scientifique d'OpenEdition, pôle qui avait notamment dans ses missions la qualité des données produites par les quatre plateformes d'Open Edition (OpenEdition Books et Journals, Hypothèses et Calenda)⁷ et leur interopérabilité⁸. Je suis donc venue à la gestion des données et à son corollaire désormais indispensable, l'*open data*, par l'*open access* et l'édition électronique⁹. C'est en raison de cette expérience professionnelle que j'ai été intégrée à ces projets par la MMSH, sans que je sois pour autant spécialiste de la question. Il n'y a en effet, à l'heure actuelle, pas d'ingénieur-e à la MMSH affecté-e expressément à des missions de gestion des données pour les projets de recherche, de manière prioritaire ou secondaire ; il faut donc faire appel soit à des personnels affectés à d'autres missions, soit aux personnels des laboratoires, elles et eux-mêmes rarement spécialisé-e-s mais se formant régulièrement via les réseaux professionnels.

Je me concentrerai ici sur l'étape préliminaire au projet européen, la rédaction de la réponse à l'appel et les questionnements qui émergent à ce

-
- 5 La partie « Communication, dissémination et exploitation des résultats de la recherche » est attendue dans le point « Mesures pour maximiser l'impact » (2.2) des projets européens. Ces trois actions sont considérées comme majeures à la fois dans la mise en œuvre du projet et dans l'évaluation de ses résultats. Elles font l'objet d'un ou de plusieurs *Workpackages* dédiés. La gestion des données, notamment la description de celles qui seront utilisées et produites, est présentée dans cette partie 2.2. L'éthique est quant à elle présentée dans la partie 4.5 des projets.
 - 6 Le *data management plan* (DMP), ou plan de gestion des données, décrit le cycle de vie des données collectées et produites dans le cadre d'un projet de recherche. Pour les projets H2020, le DMP est censé être communiqué six mois après le début de la mise en œuvre du projet, et la réponse à l'appel doit présenter dans la partie 2.2 les éléments identifiés comme pertinents pour le DMP à venir, en relation avec les impacts attendus.
 - 7 Les données produites par les plateformes d'OpenEdition sont relatives à la publication et la communication scientifiques. Si elles ne sont pas des données collectées pour la recherche – même si les données de publications peuvent faire l'objet de recherches, dans le cadre d'Open Edition Lab par exemple – la documentation numérique, le choix des formats comme des protocoles d'exposition et de diffusion de ces données, dans le cadre de l'*open access*, convergent vers les techniques mises en œuvre dans le cadre des principes FAIR et de l'ouverture des données de la recherche.
 - 8 L'interopérabilité est ici la capacité d'un système informatique à échanger automatiquement des données avec d'autres systèmes, ce qui nécessite que ces données soient présentées dans un ou des format(s) normalisé(s).
 - 9 *Open access* (aux publications scientifiques) et *open data* sont deux aspects de l'*open science*, visant à rendre les produits de la recherche accessibles au public avec le moins de restrictions possible. Pour une présentation institutionnelle et synthétique, voir par exemple « La position européenne, le mouvement de l'Open Science » : [URL : www.univ-angers.fr/fr/recherche/pour-une-science-ouverte/open-access-et-open-data/h2020-et-open-science.html].

moment-là, notamment dans la tentative de définition d'une politique des données tentant de se conformer aux exigences de la Commission européenne au sein de projets réunissant des partenaires dont les manières d'appréhender cette question diffèrent, voire divergent. Les habitudes disciplinaires, les objets de recherche ou les méthodologies habituellement pratiquées par chacun de ces partenaires, mais aussi leur environnement institutionnel et technologique, induisent un rapport assez différencié à la fois à l'ouverture des données, à l'appréhension même d'un corpus sous forme de données, et à l'éthique s'y rapportant. Mon rôle dans cette étape de montage revient à mettre en musique ces différentes manières de se poser la question des données, mise en musique qui implique évidemment ma propre éthique professionnelle.

Après une présentation succincte des deux projets et de mon rôle à cette étape de montage, j'aborderai deux exemples de dilemmes, que je qualifierais de pratique-éthiques, rencontrés pour la gestion des données, et les réponses que nous avons choisi d'y apporter.

Faire entrer les données dans les projets

Les deux projets européens objets de cet article, déposés et évalués, ont obtenu deux résultats différents : le premier n'a pas été accepté, le second a obtenu le financement demandé et a démarré en octobre 2019¹⁰.

Ces deux projets ont des dimensions différentes, autant en termes de budget, de durée que de nombre de partenaires, mais doivent tous deux répondre aux prérequis de la Commission européenne sur les données, c'est-à-dire en proposer une gestion vertueuse aux yeux de l'Europe, qu'il s'agisse d'ouverture, d'éthique ou de bonnes pratiques documentaires et technologiques.

Présentation des deux projets

Le premier projet – appelons-le « Artémis » – est une réponse au programme « Défis de sociétés. L'Europe dans un monde en évolution : sociétés

10 Pour des raisons de confidentialité, j'expose ici les éléments permettant de comprendre les questionnements que nous avons eus pendant le montage de la réponse aux appels sans entrer dans le détail des projets. Les discussions et échanges sont donc nécessairement présentés de manière simplifiée : j'espère que les porteurs et les partenaires des projets ne m'en tiendront pas rigueur. Les points de vue développés ici n'engagent que l'auteur et ne reflètent en aucun cas l'avis collectif des consortiums.

inclusives, innovantes et réflexives»¹¹, plus précisément à l'un des appels portant sur les migrations, pour un budget potentiel de trois millions d'euros sur quatre années. Artemis fait suite à un premier projet déposé deux ans auparavant dans le cadre d'un appel équivalent, projet bien évalué mais qui n'avait pas été retenu. Le consortium constitué pour Artemis réunissait onze partenaires de premier cercle, auxquels venaient s'ajouter des partenaires dits de deuxième et troisième cercles¹². Il réunissait des organismes de recherche et instituts de statistiques européens, africains et moyen-orientaux sur un sujet déjà en soi porteur d'enjeux éthiques importants, les flux et modèles migratoires, notamment vers l'Europe. Au sein de cet appel, il était expressément demandé aux consortiums de proposer des projections et scénarios sur les migrations à plusieurs dizaines d'années, afin de produire des recommandations auprès des acteurs politiques européens. Ces projections et recommandations devaient s'appuyer sur des données et des statistiques produites dans le cadre du projet : les données font donc partie des livrables scientifiques exigés, que le consortium s'engage à produire et à restituer à la Commission européenne à l'issue du projet. Le consortium avait décidé pour ce faire d'y consacrer un *workpackage*¹³ entier, donc de traiter la production des données comme une action en soi.

Le deuxième projet – appelons-le « Tanit » – est une réponse à un appel dit « Twinning » (Jumelage), qui fait partie du programme « Widespread » (Diffuser l'excellence et élargir la participation) de H2020. Ce programme vise à financer des projets qui réduisent les disparités en termes de recherche et d'innovation au sein de l'Union européenne ou entre pays de l'Union européenne et pays associés. Le projet, déposé en novembre 2018, réunit quatre partenaires, dont un issu d'un pays associé à l'Union européenne (non membre, en l'occurrence la Tunisie) ainsi que des partenaires français et belges, pour un montant d'environ 800 000 euros sur trois années. Pour ce projet, ce n'est pas tant l'objet de la recherche que la disparité entre partenaires, pourtant constitutive du programme, qui a soulevé des enjeux éthiques liés aux données et au cadre technique et légal de leur gestion.

11 Voir [URL : https://ec.europa.eu/research/participants/data/ref/h2020/wp/2018-2020/main/h2020-wp1820-societies_en.pdf].

12 Ces partenaires sont mentionnés dans la réponse à l'appel à projets mais ne reçoivent aucun financement direct de l'Europe dans le cadre du projet, et ne signent pas le *grant agreement* engageant tous les membres du consortium et la Commission européenne.

13 Définir des *workpackages* consiste à découper le projet en plusieurs grandes actions permettant de le réaliser.

La difficile intégration de la gestion des données au montage du projet

J'ai été intégrée à ces deux projets de manière quasi concomitante dans mon temps de travail, mais avec deux temporalités internes aux projets différentes : sur le premier, le plus important en termes de budget comme de nombre de partenaires, j'ai été sollicitée dès la préconfiguration du consortium ; sur le deuxième projet, j'ai été appelée à la rescousse à moins de deux mois de la date limite de dépôt de la réponse, alors que l'équipe de la MMSH s'était elle-même vu transférer la coordination du projet trois mois avant la date limite de dépôt, pour des raisons institutionnelles.

Artemis avait bénéficié de l'expérience du premier dépôt deux années auparavant : la prise de conscience de la nécessité de préconfigurer un DMP s'était faite au cours du montage de la première réponse à l'appel, dans laquelle les données avaient une place moins centrale que dans la deuxième mouture. La seule réunion rassemblant l'ensemble du consortium d'Artemis à l'étape de montage, en septembre 2018, était en partie organisée autour des données, confirmant la centralité de la question sur ce projet. Les ingénieurs en charge du développement de la plateforme de visualisation des données collectées et produites ont été invités à présenter une première ébauche de leur site à l'ensemble du consortium. Pourtant, les personnes désignées comme étant en charge de la rédaction de la partie portant sur la gestion des données au sens du DMP, dont moi-même, n'ont pas été sollicitées pour présenter l'avancement de leur réflexion aux autres partenaires et discuter avec eux. C'est le paradoxe de cette configuration qui a révélé la différence d'acceptation et de compréhension du terme *gestion de données* parmi les différents acteurs du projet.

Les discussions informelles qui en ont découlé ont confirmé que les ingénieurs en charge de la visualisation et nous, ingénieur-e-s en charge du défrichage du DMP, étions d'accord sur le fait que la plateforme de visualisation et d'exposition des données d'une part et les tâches devant être réalisées dans le cadre du DMP d'autre part – recensement, stockage, archivage, documentation, définition des modalités d'ouverture et de partage des données – ne devaient pas être confondues ni même prises en charge par les mêmes personnes ni avec les mêmes outils. Pour autant, le déroulé de cette réunion a mis au jour la confusion entre ces différentes parties du traitement des données parmi les équipes de recherche.

Les porteurs scientifiques imaginaient à cette étape de conception que les ingénieurs en charge de la visualisation des données « gèreraient » de fait les données – puisqu'ils les traiteraient –, c'est-à-dire feraient plusieurs

métiers en même temps : concevoir, développer et implémenter un système de visualisation, notamment géographique, mais aussi assurer le stockage (et sa sécurisation), l'archivage (et l'établissement des procédures d'archivage), la documentation des données, la définition et la mise en œuvre des modalités d'accès et de partage et l'éventuelle ouverture des données produites par le projet (*open data* par défaut), en prenant en compte les enjeux éthiques et juridiques qu'elles impliquent. Ce qui n'était ni dans leurs missions ni même dans leur propre projet de plateforme.

La gestion des données était à ce moment du montage du projet perçue comme une simple question de mise en application d'une demande technique de la part de la Commission européenne, un travail de traduction techno-documentaire entérinant les conclusions des discussions des équipes de recherche et intervenant en fin de course, sans dialoguer au préalable avec elles. Les incompréhensions entre les différents acteurs du projet autour des attendus de la Commission européenne en matière de gestion des données ont été révélées par les différents moments du montage. Le travail de mise en accord – probablement partiel – s'est fait par ajustements et tâtonnements d'autant plus hésitants que cette injonction est relativement nouvelle dans les projets européens, et qu'elle ne correspond ni à des besoins exprimés en ces termes par les équipes de recherche, ni à des missions et métiers clairement attribués du côté des ingénieur-e-s.

Sur le projet Tanit, j'ai été sollicitée tardivement par l'équipe en charge du montage du projet à la MMSH : d'abord mobilisé sur la conception de la réponse scientifique et organisationnelle et sur l'équilibrage fragile entre partenaires¹⁴, le consortium ne s'est attelé à la question de la gestion des données et du DMP que dans les dernières semaines avant la date limite de réponse. L'équipe qui s'est retrouvée en charge de la coordination de la réponse à l'appel, composée de la responsable scientifique et d'une ingénieure chargée des programmes européens, s'est alors rendu compte qu'elle n'avait ni les compétences ni le temps pour rédiger cette partie du projet. Elles ont fait appel à moi, identifiée par mon engagement dans Artemis, après avoir obtenu l'accord du directeur de mon unité de rattachement, ces missions ne relevant pas de ma fiche de poste initiale. Il a fallu dans ce cas penser en urgence la question des données au sein d'un projet déjà avancé, y compris dans son impact budgétaire, son architecture technologique et les compétences à mobiliser.

14 Cet équilibrage s'est traduit par le transfert tardif de la coordination du projet à la MMSH, et la nécessité d'y identifier et de mobiliser en quelques semaines les personnes compétentes sur chacun des aspects à traiter (gestion des données, éthique, gouvernance, etc.).

Si ce hors-série de *Tracés* est bâti sur l'idée d'une présence accrue de la question des données dans la recherche en sciences humaines et sociales, force est de constater que dans l'expérience concrète et la temporalité pratique du montage des réponses aux appels à projets elle arrive parfois tardivement, et reste une préoccupation, sinon annexe du moins secondaire, dans la manière dont les porteurs scientifiques conçoivent le projet.

Concilier entretiens ethnographiques et images satellites

Dans le premier projet, Artemis, la réunion du consortium puis la réception par à-coups des *workpackages* rédigés par les partenaires ont vite montré que la disparité des types de données – utilisées ou produites par les équipes de recherche – était assez forte. Ce que nous avons qualifié de *données* pour ce projet allait ainsi d'entretiens oraux avec des populations en grande vulnérabilité – voire en danger – menés sur leur lieu de transit, à des données statistiques nationales, en passant par des images satellitaires et des données d'opérateurs téléphoniques privés¹⁵, certaines de ces données devant être croisées pour mener à bien la recherche.

Derrière cette disparité se retrouvent bien sûr des différences disciplinaires ou sous-disciplinaires, et des rapports à la science qui diffèrent voire divergent, des anthropologues aux démographes spécialisé-e-s en prospective, jusqu'à l'institut chargé de proposer des scénarios pour le futur à plusieurs dizaines d'années, travaillant autant pour la recherche publique que pour des entreprises privées. L'architecture imposée de la réponse à l'appel à projets, qui découpe la recherche en différents *workpackages* portés chacun par un des partenaires, n'aide pas à ce que ces disparités se révèlent au moment de la rédaction de la partie scientifique. Ce n'est qu'une fois les différents *workpackages* rédigés que l'étendue de la problématique apparaît pour les personnes en charge des données.

Si la mise en cohérence scientifique de ces types de données n'est sûrement pas une question nouvelle, le fait qu'elle s'insère dans un livrable de type DMP et dans une injonction par défaut à l'*open data* modifie les questionnements pratiques et éthiques qu'elle pose, autant aux équipes de

15 Ces deux derniers exemples étaient évidemment ceux soulevant le plus de questions éthiques par avance : les images satellites présenteraient-elles des personnes ou groupes de personnes ? Les données téléphoniques permettraient-elles de tracer individuellement des réfugiés en Europe ? Ces dernières données devaient être fournies à un institut national de statistique, partenaire du projet, par un opérateur téléphonique privé dans le cadre d'un accord de collaboration devant être approuvé par le Comité éthique du projet.

recherche qu'à ce nouvel acteur des projets qu'est la personne ou l'équipe en charge du DMP. Il a ainsi fallu penser la gestion des données et l'architecture de la gouvernance du projet de manière à être en mesure de prendre en charge les aspects éthiques liés aux données¹⁶ (prises individuellement ou ensemble) pendant le déroulé de la recherche, tout en permettant la gestion de données de nature très différente. La place de la gestion des données a ainsi pris de l'ampleur au fur et à mesure du montage du projet, avant tout parce que traiter et documenter une telle quantité de données devenait un élément central : cela répondait d'abord à un besoin pragmatique, celui de rendre effective la fourniture de données et statistiques dans un livrable tout en balayant l'ensemble des questions de recherche des différents partenaires.

Pour ce faire, nous avons finalement décidé de prévoir l'embauche d'un-e «ingénieur-e données»¹⁷ pendant la durée du projet, en charge du DMP et de son application, dont la mission serait d'assurer l'interface entre les équipes de recherche produisant des données et les infrastructures auprès desquelles elles seraient déposées et archivées. Alors qu'il était d'abord prévu qu'un-e ingénieur-e vienne en soutien à l'équipe chargée de la plateforme de visualisation, ce poste budgétaire a été transformé pour être affecté à la MMSH, pour deux raisons : d'abord parce que cette dernière était porteuse du projet, et à ce titre le point de centralisation organisationnel et scientifique ; d'autre part parce qu'elle héberge une plateforme universitaire de données (PUD¹⁸), relais de l'infrastructure nationale de données Progedo (Production et gestion des données en sciences sociales), sur laquelle le projet devait s'appuyer pour la gestion de ses données. Cet-te ingénieur-e données aurait ainsi été intégré-e à la PUD et aurait travaillé en collaboration étroite avec son ingénieur d'études et sa référente scientifique, elle-même membre d'Artemis.

Ce choix nous a semblé nécessaire parce que le traitement et la documentation des données constituent un métier et que la compréhension des demandes des plateformes d'archivage est une compétence, et que tout cela

16 Par exemple, comment collecter, traiter, partager – y compris au sein du consortium – et exposer des données personnelles à caractère sensible, provenant de pays tiers et pouvant mettre en danger les individus ou les groupes ?

17 Nous n'avons pas discuté plus avant la dénomination du poste, même si nous savons que des intitulés concurrents ou complémentaires existent (*data scientist*, *data librarian* ou *data steward* par exemple) : le terme *ingénieur-e données* (*data engineer*) correspondait le mieux à la fois aux compétences et aux missions attendues, à ce qui était le plus immédiatement compréhensible par les différents membres du projet, et à la logique organisationnelle choisie.

18 «La PUD-AMU [...] développe une spécialisation en données quantitatives pour les études en sciences humaines et sociales dans l'aire méditerranéenne. Elle accompagne la communauté scientifique dans la recherche et l'utilisation des données quantitatives, leur traitement logiciel et leur analyse.» Voir [URL : <http://www.progedo.fr/promouvoir/plates-formes-universitaires-de-donnees/>].

prend du temps. Mais il conduit dans le même temps à renforcer la distance entre la gestion des données et leur production. Les situations concrètes d'entretien, de collecte, le rapport aux enquêtés, aux fournisseurs de données primaires, toutes ces situations qui font intrinsèquement partie de l'éthique de la recherche, se trouvent dans cette configuration « déconnectées » de l'opération de traduction en « données » pour les plateformes de gestion des données de la recherche. La mise en place de cette intermédiation – de personne et de métier – aurait produit une transformation non seulement du travail concret des chercheuses et chercheurs, qui auraient eu à penser d'emblée et chemin faisant la documentation, la transmission et l'exposition des données, mais aussi probablement de la recherche produite elle-même. Au moment de l'écriture du projet, nous n'avons pas poussé plus loin la réflexion sur ce que la gestion des données fait à la recherche, par manque de temps : ces problématiques auraient probablement pu être abordées dans l'articulation entre le travail autour du DMP et celui du comité d'éthique du projet, composé de représentants des partenaires, avec pour mission de prendre en charge les questions éthiques autour des données qui ne pouvaient qu'émerger au fur et à mesure de la réalisation du projet.

J'aurais aimé pouvoir décrire les échanges avec les différents partenaires pour articuler ces questionnements et les mettre en cohérence, dans une sorte de précomité éthique dédié aux données, mais la logistique et l'agenda du montage du projet ne les ont pas permis. Nous n'avons eu qu'une réunion à ce sujet, en visioconférence et tardivement dans le montage du projet, destinée avant tout à faire le point sur les types de données utilisées et produites. Finalement, carte blanche était laissée aux personnes en charge de la partie consacrée au DMP sur les choix techniques et documentaires autour des données, que nous avons décidé de décrire de manière assez générale en l'absence de réflexion plus poussée.

Gestion des données et souveraineté de la recherche

Le deuxième cas nous a mis face à une contradiction interne à l'appel à projets, que l'aspect éthique de la gestion des données a fait émerger. Sur le projet Tanit, le porteur doit être le partenaire du pays qui bénéficiera du projet, en l'occurrence la Tunisie. Dans les faits, le montage du projet exige une telle ingénierie – que ce soit en termes de maîtrise des attendus de la Commission européenne, de rédaction de la réponse, de montage du consortium de partenaires ou du budget – qu'il a été essentiellement réalisé par notre institution avec le partenaire tunisien, qui en est le porteur officiel. Cette

contradiction entre les exigences techniques et administratives préalables de la part de l'Union européenne et l'objet même de l'appel – monter un partenariat destiné, à l'issue du projet, à réduire les écarts entre institutions de recherche – revient à présupposer l'existence de ce que le projet est censé faire émerger, alors que l'appel mentionne comme objet même du projet les besoins en « expertise et assistance technique » et en « transfert de connaissances » identifiés auprès du partenaire du pays associé.

Cette contradiction s'est retrouvée dans la gestion des données, d'abord et avant tout parce que le niveau de structuration des infrastructures, des « bonnes pratiques », des métiers et des compétences en matière de gestion des données de la recherche était déséquilibré d'emblée. Dès les discussions pour le montage du projet, les partenaires français et belges étaient outillés pour penser et apporter des solutions clés en main (plateformes d'archivage, entrepôts de données, etc.), ancrées en Europe, à un projet où les données avaient de toute façon une place bien moindre – et a priori moins sensible – que dans le précédent. Pour les partenaires tunisiens, pour lesquels l'enjeu de la réussite de Tanit était bien plus important, la motivation centrale était avant tout de renforcer leurs liens avec des organismes européens autour de leur champ de recherche de manière à se positionner comme institution ressource. Dans la mesure où les appels Twinning ne financent pas directement la recherche, la gestion des données du projet aurait pu, pour toutes ces raisons, être assez facile à traiter du point de vue des outils techniques et documentaires à mobiliser¹⁹.

Nous nous sommes donc dans un premier temps concentrés sur les aspects légaux liés aux données, de manière à vérifier que l'environnement juridique encadrant les données de la recherche et les données personnelles en Tunisie était compatible avec le cadre européen, permettant une mise en œuvre de l'*open data* et des principes FAIR²⁰. Le cadre législatif tunisien tend à converger avec les lois européennes, notamment avec le RGPD²¹, et des discussions sont en cours pour une mise en œuvre de principes relevant de

19 Les données concernées sont soit collectées auprès d'autres opérateurs du domaine de recherche qui les mettent à disposition librement, soit collectées directement par l'équipe dans le cadre d'une enquête financée sur coûts indirects (voir note 22).

20 « Findable, accessible, interoperable and reusable » : principes de gestion des données adoptés par la Commission européenne dans le cadre du programme H2020 en 2016 qui guident les porteurs de projet de manière à permettre une meilleure ouverture et possibilité de réutilisation des données. Les principes FAIR sont exposés par Wilkinson Mark D., Dumontier Michel *et al.*, 2016, « The FAIR guiding principles for scientific data management and stewardship » [en ligne], *Scientific Data*, n° 3, [URL : <https://doi.org/10.1038/sdata.2016.18>], consulté le 4 mai 2020.

21 Règlement général sur la protection des données de l'Union européenne, texte de référence en matière de protection des données à caractère personnel.

l'open data. En l'état, il ne devait donc pas y avoir d'hétérogénéité trop forte entre les contraintes et possibilités des différents partenaires sur cet aspect.

Cependant, la contradiction entre un appel destiné à faire converger un partenaire vers des modes de fonctionnement compatibles avec les projets européens et les contraintes internes à l'appel a resurgi quand il a fallu décider où conserver les données. Nous étions partis classiquement sur un stockage et un archivage auprès d'infrastructures françaises, intégrées aux feuilles de route européennes, puisqu'elles offrent tous les services et toutes les garanties demandées par l'Europe. Or le principe même d'un projet Twinning est que les données soient collectées et hébergées avant tout par le partenaire tunisien, sur son terrain et sur le champ de recherche qu'il souhaite renforcer. Nous nous sommes alors trouvés confrontés à un dilemme que nous avons énoncé en termes de souveraineté sur les données. Au-delà de la propriété des données produites par un consortium sur financement européen, qui est une question avant tout légale et régie par le cadre réglementaire de la Commission européenne, nous avons abordé la souveraineté sur les données de ce projet comme un enjeu avant tout éthique.

Il était difficilement concevable que les données ne soient pas au moins stockées en Tunisie, de manière à ce que le partenaire tunisien puisse y accéder, les exposer, les traiter et les échanger indépendamment – autant que faire se peut – avec des infrastructures technologiques européennes, y compris au-delà de la durée du projet du consortium. Il s'agissait à la fois de diplomatie scientifique, question particulièrement prégnante pour un type de projet où l'équilibre entre partenaires d'emblée postulés comme non égaux est à construire en permanence, et de cohérence interne à l'appel puisque l'objectif est de rendre le partenaire tunisien autonome et expert sur un domaine scientifique spécifique.

Nous avons donc opté pour un système de double stockage : un stockage premier et prioritaire auprès d'infrastructures européennes, offrant toutes les garanties à la Commission européenne, et auquel évidemment le partenaire tunisien participe et a accès au même titre que les autres membres du consortium, et une réplique régulièrement mise à jour localisée chez le partenaire tunisien. Cette architecture nous a posé un deuxième problème, celui-ci très pratique mais fondamental : le partenaire tunisien ne dispose pas d'un serveur propre capable d'accueillir ces données, en termes de taille et de garantie d'autonomie ; il nous fallait donc prévoir le budget pour l'achat d'un serveur dédié. Or, les coûts d'équipement sont inéligibles aux financements des appels Twinning. Nous sommes ainsi retrouvés dans un premier temps dans l'incapacité de mettre en place l'architecture choisie pour concilier la gestion des données avec l'autonomie

souhaitée du partenaire tunisien, en raison des règles de financement de ce type d'appels²².

Nous avons donc dû nous tourner vers des structures tunisiennes déjà existantes ; mais comme il n'existe pas en Tunisie de réseau de type Renater ou d'infrastructure nationale dédiée aux données, sur lesquels pourraient s'appuyer les institutions de recherche, il a fallu faire appel à un autre acteur public tunisien, partenaire de deuxième cercle du projet mais non spécialisé dans la gestion des données, qui hébergera les données sur son propre serveur, dans le cadre d'un accord avec notre partenaire tunisien. Ce montage va permettre d'initier une coopération entre ces deux institutions tunisiennes autour de l'échange des données ; mais il laisse le porteur du projet dépendant des infrastructures européennes d'un côté et d'un acteur tunisien non spécialiste de l'autre, en matière de gestion des données qu'il aura coproduites.

Ce que les données nous disent de la recherche

De l'expérience de ces deux montages de projets et des problématiques qui ont surgi autour de la gestion des données, deux points émergent.

Le premier, qui n'est pas nouveau mais qui nécessite d'être rappelé et réinvesti pour chaque cas concret, est que la gestion des données implique plusieurs métiers et compétences, qui ne sont pas toujours identifiés par les porteurs scientifiques lors du montage de projet. De plus, les injonctions européennes (DMP, *open data*, respect du FAIR) induisent une professionnalisation de la gestion des données et une externalisation vers les infrastructures ; elles soulèvent ce faisant des enjeux éthiques sur les conditions de production et de stockage des données que le timing du montage des projets ne permet que rarement d'aborder en amont, afin d'en mesurer la portée pour chaque cas et de proposer des structures de gouvernance adaptées. Évidemment, cette prise en compte suppose que les compétences en matière de gestion des données soient déjà présentes et identifiées parmi les partenaires du consortium en cours de constitution, ce qui n'est pas toujours le cas.

Le deuxième point tient à ce que la gestion des données révèle en matière de coopération internationale dans la recherche. La question des données,

22 Une autre solution aurait été de comptabiliser cet achat dans les coûts indirects (*indirect costs*), enveloppe de 25% supplémentaires au montant du projet destinée à couvrir les frais induits pour les partenaires, mais non directement liés à la mise en œuvre du projet. Ces coûts ne nécessitant pas de justification, et la distinction entre coûts directs et coûts indirects étant parfois difficile à faire, l'enveloppe peut éventuellement être mobilisée pour des dépenses inéligibles. Dans notre cas, elle était déjà réservée à d'autres dépenses, notamment pour la collecte des données.

pourtant pensée comme une forme de rationalisation de la recherche permettant sa reproductibilité et son ouverture au bénéfice de la communauté de recherche internationale, met au jour des disparités persistantes entre pays, donc entre collègues, lorsqu'elle doit être mise en œuvre concrètement. Les principes comme les contraintes des projets européens peuvent tendre à reproduire ces disparités et à doubler la dépendance financière d'une dépendance technologique envers les infrastructures européennes.

Dans ces conditions, chercheur-e-s et ingénieur-e-s n'ont d'autre choix que de penser ensemble et en commun les questionnements scientifiques et leurs pendants techniques, parce que la gestion des données implique une éthique de la donnée qui est aussi et avant tout une éthique de la recherche.