

Page de garde		
Acronyme	PaRéDo SHS	
Titre de la proposition	Partage et Réutilisation des Données en SHS	
Nom et prénom du coordinateur ou de la coordinatrice	Dumas Primbault	Simon
Adresse email du coordinateur ou de la coordinatrice	simon.dumas-primbault@openedition.org	
Organisme du coordinateur ou de la coordinatrice	OpenEdition UAR 2504	22, rue John Maynard Keynes 13013 Marseille
Lieu de réalisation des travaux	OpenEdition, Marseille	CIS, Paris
Montant de l'aide demandée à l'ANR	298.945,76 € (frais d'environnement compris)	
Coût complet du projet	529.798 €	
Durée	36 mois (septembre 2024 - août 2027)	
Mots-clés de la proposition	données, SHS, partage, réutilisation, pratiques, recherche sur la recherche, gouvernance	
Résumé (non confidentiel) correspondant à celui enregistré sur le site de dépôt		
<p>De nombreuses études des pratiques liées aux données autant que nombre de critiques émises directement par la communauté académique montrent que l'un des verrous les plus importants au partage et à la réutilisation des données de la recherche en sciences humaines et sociales est de nature épistémologique : les principes, les politiques, les outils et, partant, le cadre de spécification, de circulation et de valorisation des données de la recherche en SHS sont susceptibles de ne pas respecter la diversité des cultures épistémiques de ces disciplines – entre autres, la diversité des matériaux, les dynamiques sociales de leur circulation, les méthodologies constructivistes qui façonnent le rapport des communautés à leurs matériaux, ainsi que l'importance de “faire” <i>son</i> terrain, ses archives ou d'entretenir un certain rapport à ses enquêtes.</p> <p>Les infrastructures de données ouvertes sont le <i>nexus</i> où s'articulent des politiques publiques, des organismes publics d'ampleur nationale et internationale, des acteurs privés, des dispositifs socio-techniques, mais aussi les pratiques matérielles d'une variété d'usagers et les données elles-mêmes. Elles sont donc un lieu de tensions productives où se joue la rencontre entre un principe d'ouverture, un besoin de standardisation – permettant le stockage, la documentation, la circulation et l'interopérabilité de données de plus en plus massives – et la nécessité d'équiper au plus juste les pratiques des SHS dans leur singularité – en France en particulier, du fait de constructions institutionnelles et infrastructurelles, et par contraste avec les STEM et leurs données – ainsi que de préserver la diversité épistémique au sein même des SHS. Afin de garantir une certaine forme d'autonomie académique (notamment au niveau des pratiques de recherche) tout en préconisant de bonnes pratiques de données ouvertes (notamment par des normes et standards), la question opérationnelle qui se pose est celle d'une gouvernance épistémique des données : comment arbitrer, en</p>		

contexte, entre la diversité des pratiques, méthodes et rapports à une multiplicité de matériaux sources, la prescription de bonnes pratiques et de standards, des impératifs techniques et des orientations politiques ?

Afin de répondre à cette question, il est nécessaire de mener à bien une recherche interdisciplinaire sur tout le cycle des pratiques de production, de mobilisation et de circulation des matériaux de la recherche dans les SHS, et de lier celles-ci à la gouvernance des infrastructures afin de comprendre comment ces pratiques tout à la fois gouvernent et sont gouvernées par les infrastructures. Au sein d'OpenEdition Lab et en partenariat avec le Centre Internet et Société, le projet PaRéDo SHS s'intéresse en profondeur aux politiques, aux dispositifs et aux pratiques de partage et de réutilisation des données en SHS à l'échelle de l'infrastructure numérique mutualisée COMMONS grâce à : (WP1) une analyse massive d'un large corpus de documents de politique relatives aux données ouvertes, de recherche sur les données en SHS et de spécifications techniques et normatives, (WP2) un pan de recherche sur la recherche sous la forme d'une ethnographie des pratiques de données des usagers de COMMONS, ainsi que (WP3) des réflexions et recommandations sur les formes de gouvernance nécessaires pour préserver la diversité épistémique des SHS et ainsi favoriser l'ouverture, le partage et la réutilisation des données.

Document scientifique

Introduction	2
Contexte	3
Des critiques de la datafication.....	3
Les données des SHS en pratique.....	4
Question de recherche	5
Terrain	6
Cadre méthodologique	7
Organisation et réalisation du projet	7
Partenaires.....	7
Work packages.....	8
Données du projet.....	12
Personnes impliquées.....	12
Calendrier.....	12
Budget – aide demandée à l'ANR.....	13
Impact et retombées du projet	13
Bibliographie	14

Introduction

Les problématiques soulevées par le triptyque données de la recherche – infrastructures – politiques remontent au moins à l'époque moderne (Wiggins et Jones, 2023). Aujourd'hui, l'adoption des techniques et outils numériques en réseau dans les pratiques de savoir soulève à nouveaux frais **la question des pratiques de partage et de réutilisation des données de la recherche, en lien avec celle de la gouvernance – tout à la fois juridique, technique et épistémique – de l'information scientifique et technique**. En complément des académies, des archives et des bibliothèques, les infrastructures de données ouvertes sont aujourd'hui les dispositifs qui résultent de l'articulation entre des politiques, des outils, des standards, des usages ainsi que les données même, afin de soutenir et d'orienter la recherche scientifique. En ce *nexus* infrastructurel se jouent des relations de pouvoir asymétriques entre toutes les parties prenantes – institutions publiques, organismes nationaux et internationaux, acteurs privés, personnel, usagers – qui font et refont les dispositifs socio-techniques au gré des politiques publiques, des marchés, des développements techniques ou des cadres épistémologiques. Au sein de la gouvernance des infrastructures, **le rapport aux pratiques de recherche en SHS est donc dialectique** : en équipant la recherche, **les infrastructures ont une forme de pouvoir dans la gouvernance des SHS** en prescrivant les bonnes pratiques, en favorisant certains objets et certaines méthodes au détriment

d'autres, en transformant la nature de l'information manipulée et donc les opérations de traitement de l'information qui président à la fabrique des connaissances ; **réciroquement, la présence active d'usagers et praticiens dans les infrastructures**, leurs observatoires, conseils et instances décisionnelles – à diverses échelles, des décisions quotidiennes aux stratégies de plus long terme – leur permet dans une certaine mesure de **participer de la gouvernance afin que l'infrastructure équipe les pratiques au plus juste et incarne des valeurs partagées**. On entend donc le terme de gouvernance tant comme les formes de gouvernance *des* infrastructures par ses parties prenantes, que comme la gouvernance *par* les infrastructures des pratiques en leur sein. Dans le contexte du mouvement *open data*, c'est cette gouvernance dialectique entre infrastructures de données ouvertes et pratiques de partage et de réutilisation en SHS que le projet PaRéDo SHS se propose de mettre en lumière.

Contexte

Des critiques de la datafication

Au sein des SHS, certains principes, outils ou politiques de partage et réutilisation des données de recherche ont pu être reçus avec scepticisme par certaines communautés de recherche. De manière générale, les critiques émises par les communautés académiques à l'encontre des recommandations, incitations ou injonctions aux données pointent au moins deux éléments d'un processus de "datafication" :

1. la transformation des matériaux de la recherche en données – et modification en conséquence des pratiques de gestion, manipulation, circulation des matériaux comme données ;
2. la progression d'un paradigme épistémologique d'ordre computationnel – qu'il soit complémentaire des approches dites "traditionnelles" ou entendu comme opération de naturalisation des SHS ;

Concomitant des politiques de données ouvertes (Kitchin, 2022 ; Goëta, 2024), les débats sur la datafication des SHS remontent au moins à l'émergence des humanités numériques au seuil du XXI^e siècle (Mounier 2015, 2018). Le "tournant computationnel" (Mabi *et al.*, 2017) a pu être pratiqué comme une entreprise de transformation épistémologique des SHS visant à rendre celles-ci "*data-driven*", c'est-à-dire répondant à des impératifs épistémiques de formalisme, d'exhaustivité, de calculabilité et de reproductibilité venus des sciences des données et de certaines sciences naturelles (Strasser, 2012). Le postulat de cette transformation étant qu'une des vertus du tournant computationnel serait de permettre d'explicitier, donc de "désambiguïser", l'activité scientifique, entendue comme processus de récolte et d'analyse de données.

Ce programme épistémologique de datafication prêté au tournant computationnel par ses critiques entrerait en contradiction avec, d'une part, la singularité épistémique des SHS – leur objet général, leur démarche herméneutique plutôt qu'explicative, leur réflexivité critique – et, d'autre part, les diversités épistémiques internes aux disciplines qui composent les SHS – les échelles d'étude de l'objet, la nature des matériaux construits en conséquence, le spectre de méthodes allant du qualitatif au quantitatif.

À titre d'exemple, la lettre bimensuelle de l'Institut des sciences humaines et sociales du CNRS illustre bien la diversité autant que la singularité qui rassemble des démarches de recherche aussi variées que celle d'un anthropologue qui a passé plusieurs années sur un point de trafic de drogue à Chicago ou d'une ethnologue qui étudie les rites de passage à l'âge adulte en Inde du Sud, d'un philosophe exégète du second Heidegger, d'une historienne qui étudie la correspondance de Flaubert, d'une sociologue en STS qui déploie des méthodes computationnelles sur un réseau social, d'un archéologue qui exhume des artefacts d'une civilisation éteinte sur un chantier de fouilles, d'une professeure en littérature qui pratique la recherche-crédation avec ses étudiants ou encore d'une psychologue qui reproduit une expérience de psychologie sociale sur le lobe frontal...

Ainsi, à l'univocité de l'analyse de données il est opposé la continuelle réinterprétation de matériaux, à la réification des données et leur distinction franche de la publication il est opposé la nécessité de construire ses données, en contexte, en tenant compte d'une méthode (ex. "faire son terrain" en anthropologie, "le goût de l'archive" en histoire), à l'objectivité algorithmique des résultats de recherche il est opposé la coexistence de savoirs situés (Gitelman, 2013 ; Leonelli, 2015 ; Bonde Thylstrup *et al.*, 2019).

Plus récemment, les réflexions d'Erzsébet Tóth-Czifra, ancrées dans son expérience, dans certaines études de pratiques (notamment Horsley et Priddy, 2018) et dans l'analyse d'infrastructures existantes, alertent sur les potentiels effets néfastes de la FAIRisation des données en SHS. Selon cette ancienne *open science officer* de DARIAH-EU, si les spécificités épistémiques de ces dernières ne sont pas prises en compte, le risque n'est pas

seulement d'invisibiliser tout ce qui n'a pas été numérisé (Bardiot *et al.*, 2022), de méconnaître le décalage entre un matériau et sa remédiation numérique mais encore de perdre la *“thick description”* chère aux SHS les plus qualitatives. Pour Tóth-Czifra en 2020, les normes et infrastructures n'étaient pas encore en adéquation avec ces cultures épistémiques : *“As a result, the actual continuum of the knowledge creation procedures of the cultural heritage domain is barely reflected in its infrastructure and data management practices.”* (Tóth-Czifra, 2020, p. 249)

Ainsi, **les recommandations, incitations ou injonctions aux données** – qu'elles soient politiques, techniques ou épistémologiques – **peuvent-elles s'exercer comme une forme de gouvernance épistémique justifiée en raison technique par l'infrastructure** – qui plus est une gouvernance exogène à la communauté épistémique, résultant de l'asymétrie entre toutes les parties prenantes des infrastructures et donc une atteinte aux libertés académiques, traditionnellement assurée par la collégialité (Musselin, 2021).

Les données des SHS en pratique

En-deçà de ces critiques, souvent vives, de la datafication, de nombreuses études de terrain ont permis de **documenter plus en détail les pratiques relatives au partage et à la réutilisation des données en SHS et d'identifier les freins à celles-ci**. Si les freins juridiques – propriété intellectuelle, anonymisation – ou les freins techniques – architectures logicielles, standards – sont souvent mis en avant, les freins épistémiques sont moins souvent discutés.

En France, dès 2014, une étude de Francesca Cabrera montre l'irréductible diversité de ce que les chercheurs en SHS appellent plus spontanément, en fonction des disciplines, matériaux, sources, notes, entretiens... et souligne également des distinctions entre *“données sources”* et *“données produites”*, attestant d'un chronophage travail de *“brutification”* et de documentation des données. Cette diversité de natures et de méthodes autant que les contextes de production des données et le rapport très personnel des chercheurs à celles-ci expliquent pourquoi, malgré une volonté annoncée de partager et réutiliser leurs données, les chercheurs en SHS n'avaient alors que peu recours à ces pratiques. L'étude encourageait ainsi à accompagner les politiques de *“propositions provenant des propres communautés scientifiques, visant à combler les écarts techniques, technologiques et culturels avec un transfert de technologies et de méthodologies, instaurant ainsi une dynamique de coopération entre les différents domaines des SHS”* (Cabrera, 2014, p. 107).

En 2016 et 2017, deux études successives à l'Université de Lille 3 (Prost et Schöpfel, 2016) et l'Université de Rennes 2 (Serres *et al.*, 2017) attestent des mêmes difficultés, insistant à nouveau sur la diversité des matériaux, de leurs supports et des pratiques, autant que sur la difficulté à distinguer données sources, données produites et publications. Cependant, **grâce à des entretiens semi-directifs** qui complètent les questionnaires formatés, ce sont les **freins épistémologiques** qui ont la part belle dans l'étude de Serres *et al.* (2017). Ce que met en avant le rapport c'est une **position constructiviste**, largement partagée dans les SHS – dans le cas de l'étude, au moins en linguistique, sociologie, histoire, géographie et sciences de l'information et de la communication – selon laquelle il n'existe pas de données *“brutes”* : il n'existe que des *“obtenues”*, construites dans un contexte d'enquête toujours singulier, grâce à des outils et des techniques d'observation, de collecte, de production de données qui ne sont jamais neutres mais sont matériellement et culturellement situés. Il en découle un impératif épistémique – celui de ne pas réduire et décontextualiser les données – et un impératif éthique – reconnaître et valoriser le travail de construction.

Il y aurait donc un **paradoxe majeur** – Bendjaballah et Garcia (2023) parlent aussi d'*“injonctions contradictoires”* – **niché au cœur de la question des données : afin de partager et réutiliser des données, il faudrait transformer celles-ci en inscriptions** – au sens latourien d'écritures mobiles et immuables – **ce qui représente non seulement un travail conséquent mais, plus encore, une opération en contradiction avec la posture constructiviste**. En effet, les données sont une catégorie relationnelle (Leonelli, 2015), c'est-à-dire que ce qui fait donnée dépend non seulement de la nature de l'objet, du moment de son cycle de vie mais également des modalités pratiques de saisie de cet objet par les usagers – par exemple, selon les régimes de navigation, le corpus de Gallica peut-être entendu comme un ensemble ordonné d'œuvres, une base de données bibliographiques ou une masse d'information (Dumas Primbault, 2023a ; Kaabachi et Dumas Primbault, 2023). Le rapport conclut ainsi à *“l'impossibilité d'une politique générale, unique et identique pour tous”* (Serres *et al.*, 2017, p. 118) et préconise de *“rejeter tout discours d'injonction, de pression moralisatrice ou culpabilisante sur la question du partage des données”* (p. 119).

Plus récemment, l'étude dirigée par Célya Gruson-Daniel (Inno3) pour le Comité pour la science ouverte (Gruson-Daniel *et al.*, 2022a) montre encore, toutes disciplines confondues, que les pratiques de partage se

restreignent bien souvent à des échanges pair à pair et que la réutilisation des données est encore assez limitée. Le rapport final enjoint à “appréhender la diversité des pratiques” selon trois axes : 1) les “démarches de recherche” – *i.e.* les “principes épistémologiques, méthodologiques, critères de scientificité, paradigmes de recherche, communautés épistémiques” – ; 2) les “pratiques”, associées à “la notion de communautés de pratiques et d’apprentissages” ; et 3) les “statuts” et les “fonctions” des usagers des données.

En Amérique du Nord, d’autres études (Akers et Doty, 2013 ; Thoergensen, 2018) mettent en lumière les mêmes difficultés à nommer “données” les matériaux, sources, notes, entretiens allant jusqu’à qualifier l’usage du mot “*data*” dans certains contextes de “*dirty word*” (Mohr *et al.*, 2016). Pour Mostern et Arksey (2016) dans un article intitulé “*Don’t Just Build It, They Probably Won’t Come*”, c’est de la méconnaissance de ces spécificités épistémiques – même au sein de sciences sociales quantitatives – que résulte une série d’échecs de plateformes dédiées aux données historiques qui n’ont pas su prendre en compte la “vie sociale des données”, c’est-à-dire leur aspect construit au sein de communautés de pratiques.

Ainsi, l’écart entre les normes de bonnes pratiques et les pratiques réelles, et donc l’hétérogénéité de ces pratiques, ne doit pas systématiquement conclure à la non-conformité desdites pratiques en regard des principes de la science ouverte mais permettre, réciproquement, de repenser ceux-ci. En conséquence, **afin que les dispositifs socio-techniques n’agissent pas seulement sur les acteurs des données comme une contrainte épistémique exogène à la communauté académique mais que ceux-ci puissent en retour participer à la définition des normes, il est nécessaire que la gouvernance des données au sein des infrastructures reflète les spécificités épistémiques des SHS.**

À noter que le risque que font courir certains principes de science ouverte sur la diversité épistémique des STM a également été mis en évidence par Sabina Leonelli (2022, 2023). Ainsi, par exemple, le critère de reproductibilité n’a-t-il pas la même signification dans tous les champs et n’est pas nécessairement garant de la qualité d’une recherche (Leonelli, 2018). Cependant, si la problématique générale est similaire au sein des STM, les spécificités des SHS – les asymétries de pouvoir et de financement souvent en défaveur de ces dernières, l’origine de ces principes qui émanent de certaines disciplines des STM, le rapport singulier des SHS à leurs matériaux, le rapport de gouvernance dialectique entre pratiques et infrastructures plus marqué dans les SHS – font de celles-ci le meilleur lieu d’observation de ces tensions.

Question de recherche

Afin de mieux répondre aux exigences épistémiques des SHS en matière de données sans céder à une critique abstraite de l’*open data* comme datafication forcée, l’ambition de ce projet est de **mener une analyse des effets des infrastructures de données ouvertes sur les pratiques de recherche en SHS et réciproquement**. En effet, la gouvernance dialectique entre infrastructures et pratiques est une forme de “gouvernance épistémique” (Vadrot, 2011) – entendue comme “*the relationship between policy making and knowledge production through research*” (Lund *et al.*, 2022, p. 536). Lund *et al.* sont particulièrement attentives aux effets délétères de certaines formes de gouvernance qui peuvent produire des injustices épistémiques (Visvanathan 1997 ; Fricker, 2007) et soulignent le rôle des SHS pour comprendre ces mécanismes et préserver la diversité épistémique.

Ce qui se joue dans la problématique de la diversité épistémique au sein des infrastructures de données, c’est **une tension infrastructurelle essentielle entre standardisation et flexibilité, redoublée d’une tension entre constructivisme et réalisme des données**. Une tension encadrée et rendue productive par des formes de gouvernance particulières. En effet, Fisher et Streinz (2022) ont montré que les “inégalités de données [*data inequalities*]” sont fonction du “contrôle infrastructurel [*infrastructural control*]” exercé sur et par les données et les pratiques, nécessitant de développer une gouvernance commune, en d’autres termes des communs de la connaissance (Frischmann *et al.*, 2014, Hess et Ostrom, 2007). Dans le cas des données, cependant : “*Governance models for open access to data are nascent at best [...] The essential questions to address about any commons concern equity, efficiency, and sustainability*” (Borgman, 2020).

Les infrastructures numériques de savoir sont le lieu privilégié de la négociation et de l’exercice d’une telle gouvernance épistémique car elles sont devenues le principal maillon d’articulation, *via* des dispositifs socio-techniques, entre la diversité des pratiques et les normes de bonne pratique. Elles peuvent être aussi bien le véhicule d’un programme computationnel que le moyen d’équiper de façon discriminante les épistémologies spécifiques aux SHS. En effet, prenant de plus en plus l’aspect de plateformes (Plantin et Thomer, 2023), elles se

situent en des points névralgiques du cycle de vie des objets autant que des réseaux de circulation des savoirs et de l'information. Loin d'être des médiatrices neutres, les infrastructures influent sur les objets qui y circulent, sur les pratiques associées à ces objets, autant qu'elles structurent localement l'"écosystème" dans lequel elles s'insèrent. En retour, elles sont gouvernées par les diverses parties prenantes, en particulier leurs usagers, autant qu'elles s'adaptent à des sollicitations émanant de leur position dans l'écosystème.

Selon notre modèle écologique (Mounier et Dumas Primbault, 2023a), la gouvernance des infrastructures à la fois cadre et subit l'interaction dans des "trading zones" d'une grande diversité de "stakeholders" afin d'assurer que l'infrastructure incarne techniquement un ensemble d'objectifs négociés par les parties prenantes – par exemple durabilité, transparence, ouverture, inclusion, diversité. **Le point de départ du présent projet est de reconnaître que, d'une part, cette gouvernance a un effet sur les pratiques de recherche et qu'en retour elle est sous influence de ces pratiques.** Ainsi, la gouvernance épistémique des données, résultant de cette dialectique, en recueillant la diversité des pratiques et en préconisant et légitimant de bonnes pratiques incarnées dans des dispositifs infrastructurels de tous ordres, met en jeu d'une manière particulière la question de l'autonomie scientifique. Il s'agit donc de qualifier en contexte ce lien dans les infrastructures de données afin de **préconiser des formes de gouvernance préservant la diversité épistémique des SHS.**

En-deçà ou en amont des données, ce projet propose de **revenir à la diversité formelle et matérielle des matériaux mobilisés par les différents champs et acteurs des SHS.** À partir de cette diversité, il sera possible de **documenter, par les pratiques, la représentation numérique de ces matériaux**, qu'ils aient été numérisés ou soient nativement numériques, **et comment celle-ci façonne ou est façonnée par les pratiques de recherche** selon les objets, les disciplines et les matériaux. Ainsi, le rapport des chercheurs en SHS à leurs matériaux à l'ère du numérique sera documenté sur tout leur cycle de vie, depuis leur production, le cas échéant, leur numérisation, leur analyse ou interprétation, leur mobilisation dans des publications, leur stockage et conservation à court terme en local, puis à moyen et long terme dans des dépôts, puis leur partage et leur réutilisation.

Prises dans leurs liens avec l'infrastructure, il s'agira alors de **documenter comment ces pratiques de recherche en SHS sont gouvernées par l'infrastructure** – standards et formats, recommandations de bonnes pratiques, modalités d'accès, visibilité et légitimité – **et quel est l'effet de cette gouvernance sur la diversité épistémique ; réciproquement, il sera possible d'observer comment ces pratiques** – par des mécanismes d'observatoires, d'instances décisionnelles, de co-construction – **gouvernent et orientent l'infrastructure dans son développement afin qu'elle équipe la recherche au plus juste en respectant la diversité épistémique des SHS.**

Terrain

Le recul dont nous disposons aujourd'hui sur les politiques infrastructurelles de données permet désormais de mener une étude de terrain d'ampleur sur les effets de gouvernance épistémique réels des infrastructures de données sur les pratiques de recherche en SHS et réciproquement. Il est toutefois important de se pencher sur ces problématiques avant que les infrastructures n'atteignent une échelle générant des dépendances au sentier contre l'inertie desquelles il devient difficile d'intervenir. Du fait de sa relative jeunesse, de sa taille maîtrisable autant que de la richesse des usages qu'elle rend possible et de son importance dans le champ des SHS francophones, **l'infrastructure numérique mutualisée COMMONS, dont le développement est coordonné par OpenEdition**, en étroite collaboration avec Huma-Num et Métopes, représente une étude de cas unique.

De plus, **tout le cycle de vie des matériaux**, données et publications est couvert par les services de COMMONS : de la collecte et du traitement des matériaux – Huma-Num Box, hébergement, machines virtuelles – aux publications et leurs usages – Métopes et OpenEdition – en passant par l'exposition et la conservation des données – Nakala –, ainsi que leur réutilisation – Isidore. Par ailleurs, COMMONS est une infrastructure mutualisée dédiée à **tous les champs et toutes les disciplines des SHS.** Les consortiums-HN seront particulièrement intéressants du point de vue de la diversité épistémique car ils forment des communautés scientifiques rassemblant des professions et compétences variées – chercheurs, ingénieurs, archivistes... – qui travaillent sur des objets communs dans des perspectives interdisciplinaires structurées par des méthodes et outils partagés sur des objets communs.

Cadre méthodologique

Dans le cadre conceptuel d'une écologie de la gouvernance des infrastructures (Mounier et Dumas Primbault, 2023a) et dans la continuité de l'observatoire des usages du projet COMMONS (OE Lab) ainsi que de l'analyse des voies d'impact économique et sociétal de la science ouverte du projet PathOS (CIS), **ce projet de "recherche sur la recherche" se devra d'être fortement marqué par l'interdisciplinarité**. Il sera nécessaire de tisser ensemble une variété de points de vue théoriques dans un chassé-croisé méthodologique :

- les *science and technology studies* (STS) permettront de comprendre les infrastructures de données ouvertes comme des dispositifs socio-techniques qui prennent place dans un écosystème plus large ;
- les sciences de l'information et de la communication (SIC) mettront en lumière comment, au contact de ces infrastructures, les usagers déploient des pratiques d'appropriation, contournement, détournement afin de satisfaire leurs besoins d'accès, de partage et de réutilisation de données ;
- la science des données permettra d'exploiter les traces laissées par les pratiques, grâce à des outils permettant d'identifier des régimes d'engagement entre données et entre plateformes.

Ce chassé-croisé interdisciplinaire permettra d'articuler sur plusieurs échelles d'analyse :

- **une approche qualitative** basée sur des méthodes d'enquête issues de l'ethnographie (entretiens, questionnaires, observations, groupes de discussion) ainsi que de la sémiotique (étude des interfaces, des classifications, des outils de recherche et hiérarchisation du contenu) ;
- avec **une approche quantitative** qui mobilisera de la fouille de texte sur un corpus massif ainsi que des méthodes de *machine learning* sur les traces d'usages (logs serveurs, données Matomo, *web crawling*) afin d'identifier des motifs récurrents susceptibles d'indiquer des typologies d'usages et d'utilisateurs.

Organisation et réalisation du projet

Partenaires

Le projet PaRéDo SHS verra la collaboration entre :

OpenEdition Lab, service de l'unité d'appui et de recherche OpenEdition, dont la mission est de définir, de soutenir et d'organiser, au sein de et avec l'unité, un programme de recherche interdisciplinaire, à la fois fondamentale et appliquée, sur les pratiques de production, circulation, appropriation et régulation des savoirs de SHS en régime de science ouverte. Au sein d'OE Lab, le projet s'appuiera sur l'Observatoire des usages (COMMONS), le groupe de réflexion sur la gouvernance des infrastructures numériques de savoir (GDKI) et le cycle de rencontres autour des Savoirs Partagés organisé conjointement avec le CIS et Inno3 (SaPa).

Le **Centre Internet et Société**, qui entend construire une recherche et une expertise indépendantes et interdisciplinaires, capables d'éclairer les grandes controverses techniques et la définition des politiques contemporaines liées au numérique, à l'Internet, et plus largement à l'informatique. Ce projet s'appuiera sur les problématiques traitées dans les deux projets suivants portés par le CIS et dont il pourra réutiliser les acquis : l'étude des voies d'impact économique et sociétal de la science ouverte par le projet européen PathOS ainsi que l'analyse juridique des écosystèmes de données ouvertes dans le cadre des PhDs de l'ITN ODECO.

Deux recrutements seront réalisés pour mener à bien ce projet : **1 postdoctorant en sciences de l'information et de la communication chez OpenEdition**, avec une bonne connaissance du champ des données en SHS et **1 ingénieur de recherche au CIS**, avec de l'expérience dans les méthodes ethnographiques qualitatives (entretiens observations) et préférablement quelques méthodes quantitatives (analyse de questionnaires, fouille de texte, analyse de logs). Ils collaboreront sur une durée de 24 mois et sur chacun des WP. Ils seront assistés par d'autres collaborateurs du projet là où cela est mentionné.

Work packages

– **WP1 Analyse massive des littératures sur les données** : état des lieux des littératures existantes – tous genres confondus : essais, études, prises de positions, recommandations, politiques, chartes, documentation

technique... – sur les outils, les pratiques, les politiques de données numériques en SHS depuis l'apparition des infrastructures numériques de données dans le courant des années 2000.

Objectif : À la suite de la littérature déjà analysée dans le cadre des recherches menées par les partenaires de ce projet, cette généalogie transverse à toute la littérature impliquée permettra de dresser un contexte à l'analyse des pratiques, depuis l'émergence des problématiques liées aux données numériques en SHS et la création des premières infrastructures dédiées, les critiques émises à cet égard ainsi que la façon dont les politiques ont réagi à cette dynamique. Une telle analyse massive de toutes ces littératures – quantitative et computationnelle, faisant appel à des méthodes d'analyse de réseaux, de sémiotique ou d'analyse de discours – permettra de cartographier en détail les liens entre principes, politiques, infrastructures, critiques, pratiques et communautés.

T1.1 Constitution d'un large corpus et définition d'une méthodologie de type PRISMA “*scoping review*” – afin qu'il soit interopérable avec l'analyse du WP1 de PathOS. Un ensemble de plateformes (OE, Cairn), d'agrégateurs de littérature scientifique (WoS, Scopus) et de documents de politique (PolicyCommons) seront moissonnés sur la base de mots-clés choisis afin de constituer un corpus de documents textuels traitant explicitement des données en SHS depuis 2000.

1. Grâce à des outils *ad hoc* de fouille de texte et de détection de communautés développés en Python, il s'agira de dresser un panorama extensif des genres discursifs, des prises de position en fonction des postures d'énonciation et des communautés. Cela pourra être réalisé en lien avec le doctorant en informatique de l'Observatoire des usages d'OE Lab.
2. Il sera extrait de chaque panorama quantitatif un ensemble de documents représentatifs des études et postures identifiées permettant de qualifier plus précisément les questionnements, critiques et débats. L'étude qualitative de ces documents et de leurs liens est spécifiée dans les trois tâches suivantes.

T1.2 Revue de littérature émanant des SHS : par le biais d'essais réflexifs, d'études de terrain ou parfois de prises de position quasi-pamphlétaires, les SHS se sont tôt penchées sur la question de la nature et du rôle des données numériques au sein de leur singularité épistémique et en lien avec les infrastructures dédiées. Cette tâche permettra de mettre en lumière, de façon diachronique et par communautés, les observations et prises de positions, en faveur ou en opposition, sur les données au sein même des SHS. Une part importante de cette tâche a été réalisée en préparation de ce projet et dans le cadre des recherches respectives des partenaires.

T1.3 Analyse des documents de politique, de stratégie, d'orientation : cette tâche se concentrera sur l'analyse de documents de politique, de stratégie ou d'orientation émis à différentes échelles par différentes institutions – de l'UNESCO aux communautés scientifiques, en passant par la Commission européenne, la loi numérique, les PNSO et les chartes d'universités. À la suite d'Albornoz *et al.* (2018) au sujet des documents de politique de science ouverte, il s'agira de retracer les valeurs et les dynamiques portées par les institutions émettrices, les modalités de performance de ces documents – par ex., de la recommandation à l'injonction – et leurs effets potentiels en termes de gouvernance épistémique. Cette tâche pourra être menée en partenariat avec le postdoctorant COMMONS sur l'évaluation des politiques publiques.

T1.4 Analyse de la documentation technique / sémiotique de plateformes : cette tâche permettra de déterminer de quelle manière les deux tâches précédentes s'articulent et s'incarnent dans l'infrastructure. En se limitant cette fois-ci à COMMONS et en se penchant à la fois sur sa documentation socio-technique – dépôt de projet, cahier des charges, plan de gestion des données, organigramme – et sur la sémiotique de ses outils et plateformes – architectures logicielles, systèmes d'informations, interfaces –, il s'agira de documenter comment les négociations entre toutes les parties prenantes au sein des infrastructures s'incarnent dans des dispositifs socio-techniques qui équipent et orientent les pratiques et sont, en retour, orientés par les pratiques. La doctorante en SIC d'OE Lab apportera son expertise sur les dispositifs.

Livrables : D1.1 Jeu de données bibliographique (corpus)
D1.2 Rapport d'étape
D1.3 Exposition des méthodes quantitatives en *open source* sur le git d'OpenEdition Lab

– **WP2 Lier pratiques et infrastructures de données en SHS** : déployer à l'échelle de COMMONS diverses méthodes d'ethnographie des pratiques de recherche afin de documenter en détail, toutes disciplines et tous métiers confondus et tout au long du cycle de vie, les pratiques relatives aux données dans les SHS – et lier celles-ci à l'infrastructure dans sa dimension technique et politique.

Objectif : Faisant fond sur le contexte établi dans le WP1, ce bloc de recherche sur la recherche, central tant chronologiquement que scientifiquement, entend mettre à jour et prolonger les études de Cabrera (2014), Prost et Schöpfel (2016), Serres *et al.* (2017), Gruson-Daniel *et al.*, (2022a) grâce au terrain qu'offre COMMONS. Un chassé-croisé méthodologique d'approches qualitatives et quantitatives – tirées de la sociologie, de l'ethnographie, de la sémiotique, de la science des données ou encore des humanités numériques – permettra de mettre en évidence les freins épistémologiques au partage et à la réutilisation des données, ainsi que de les envisager selon trois perspectives complémentaires : les usagers (T2.1), les données (T2.2) et les publications (T2.3).

T2.1 Étude des pratiques de données des usagers de COMMONS, toutes disciplines et tous métiers confondus, grâce au tissage d'un ensemble de méthodes qualitatives, quantitatives et computationnelles :

1. un **questionnaire général** adressé à tous les usagers de COMMONS (à paragonner sur l'étude n-cliq 2022 chez OE), centré sur les données, leurs pratiques et leurs freins épistémiques ;
2. une cinquantaine d'**entretiens semi-directifs** avec un ensemble d'enquêtés représentatifs de la diversité épistémique des SHS, les entretiens seront transcrits grâce à un whisper dockérisé et codés grâce à RQDA afin de les rendre interopérables avec ceux de l'Observatoire des usages d'OE Lab ;
3. des **observations in situ** (ex. consortiums-HN) ou **en ligne** (partage d'écran et tâche à réaliser) ;
4. l'**analyse de logs** à l'échelle de COMMONS pour documenter la circulation entre les diverses plateformes, en lien avec le doctorant en informatique de l'Observatoire des usages.

Cette tâche permettra de mettre en évidence autant la singularité épistémique des SHS, que diverses cultures épistémiques internes à celles-ci et de documenter la remédiation numérique de pratiques et méthodes dites "traditionnelles", d'identifier l'émergence de nouveaux objets et nouvelles pratiques, de mettre en lumière toute l'infrastructure humaine et les compétences distribuées nécessaires à la gestion des données et l'influence de celle-ci sur la diversité épistémique. Enfin, cela permettra d'objectiver le rapport qu'ont les usagers aux principes et politiques de données, tant dans leurs pratiques que dans leurs discours – on pourra par exemple, orienter certains entretiens sur l'anonymisation des données en lien avec le projet ETAPSS coordonné par Guillaume Garcia – ainsi que leur relation aux infrastructures de données, *via* leur gouvernance.

À titre d'exemple, la figure 1 ci-dessous illustre le chassé-croisé méthodologique, entre analyse de logs serveurs et entretiens semi-directifs, mobilisé dans le cadre d'une ethnographie des pratiques de navigation d'usagers de Gallica (Dumas Primbault, 2023a). Il ne s'agit pas d'une modélisation de l'ensemble de la démarche portée dans le projet ANR dans son ensemble qui devra également articuler le questionnaire et les observations, ainsi que les tâches suivantes.

T2.2 Biographies de jeux de données : réaliser la biographie, sur tout leur cycle de vie (production, traitement, publication, conservation, réutilisation...), d'un petit nombre de jeux de données – ou ensemble de matériaux – représentatifs de la diversité des formats, des méthodes, des pratiques matérielles et des circulations en SHS. Cette perspective sur les données mêmes, s'inspirant des "*data journeys*" de Leonelli et Tempini (2020), s'appuiera, par exemple, sur un corpus de sources historiques de l'archive et sa numérisation à la constitution d'un corpus, son interprétation, sa mobilisation comme élément de preuve dans une publication et sa réutilisation dans d'autres travaux (Bardiot *et al.*, 2022) ; un ensemble hétérogène de carnets de notes, de photographies, d'enregistrements audios résultant d'un terrain anthropologique à anonymiser pour publication ; ou bien des logs serveurs produits pour l'entretien des infrastructures et détournés par des pratiques d'inscriptions et de computation afin de renseigner sur les usages (Dumas Primbault, 2024).

T2.3 Data papers et data journals : poursuivre l'étude des pratiques en adoptant la perspective des publications que sont les *data papers* et *data journals* ou, plus généralement en SHS, les formats et espaces de publications dédiés à la valorisation des matériaux de la recherche. En partenariat avec l'élément 3 du projet COMMONS sur le lien données-publications, cette tâche entend s'intéresser plus avant à ces pratiques (Féret *et al.*, 2021 ; Guigonis *et al.*, 2023) ainsi qu'à des revues en transition vers OpenEdition (ex. *Sources*) ou à des

initiatives éditoriales émergentes (ex. *Cambouis*). Cette tâche permettra de mettre en évidence comment ces formats de publication de données influent sur les pratiques – ex., le *peer-reviewing* de données (Baillot 2016).

Livrables : D2.1 Exposition des données quantitatives et qualitatives des enquêtes
D2.2 Exposition des méthodes quantitatives en *open source* sur le git d'OpenEdition Lab
D2.3 Rapport d'étape

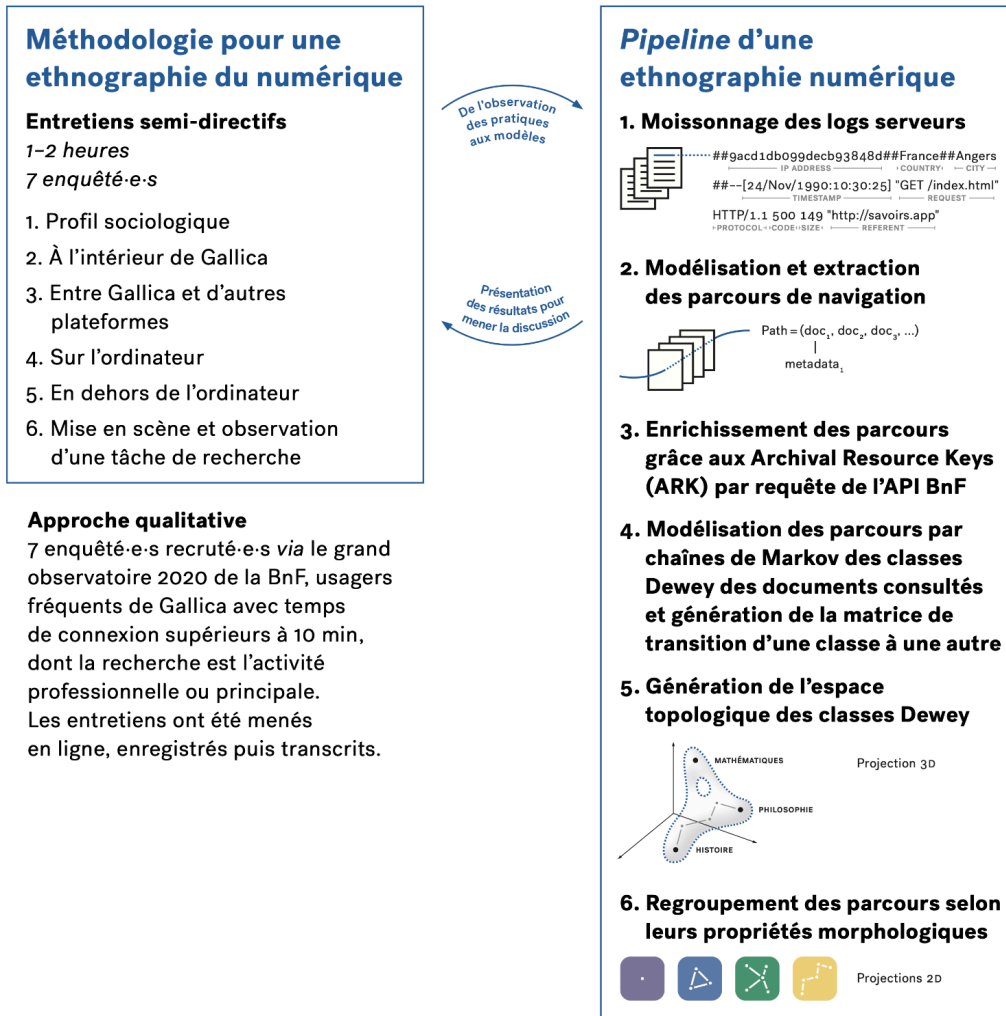


Figure 1. Exemple d'un chassé-croisé méthodologique mobilisé dans une ethnographie des pratiques de navigation sur Gallica (Dumas Primbault, 2023a).

– **WP3 Étude de la gouvernance épistémique des infrastructures de données ouvertes** et en particulier concernant les pratiques de partage et réutilisation des données en SHS.

Objectif : Faisant fond sur le séminaire “*Governing Digital Knowledge Infrastructures*” (GDKI) organisé en 2021-2022 à l’EPFL par Pierre Mounier et Simon Dumas Primbault (2023b, 2023c), ce bloc de travail se propose de mobiliser le champ des *infrastructure studies* – qui s’étend des STS aux sciences politiques en passant par le droit – afin de mieux comprendre, par le rapprochement d’études de cas réalisées par nos collaborateurs du réseau GDKI, les diverses formes de gouvernance des infrastructures numériques de savoir. Cette prise de recul comparative, parallèle au bloc précédent, permettra de mieux saisir les enjeux de la gouvernance épistémique des données qui se manifestent dans les pratiques.

T3.1 Séminaire GDKI : organisation d’un séminaire mensuel qui réunira les collègues du réseau GDKI, notamment Chérifa Boukacem-Zeghmouri (PU Université Lyon 1), Maciej Maryl (Académie polonaise des sciences), Ramya Chandrasekhar (CIS, ODECO), Bhargav Srinivasa Desikan (Institute for Public Policy Research, Londres), Suzanne Dumouchel (CNRS), Marcel Wrzesinski (Alexander von Humboldt Institut für Internet und Gesellschaft). Au cours de ce séminaire seront abordées diverses études de cas (ex. OpenEdition, EOSC, Wikipédia ou Sci-hub).

T3.2 Formes de gouvernance : formuler des recommandations à l’égard des formes de gouvernance les plus susceptibles, au sein des infrastructures de données ouvertes, de garantir la diversité épistémique.

Livrables : D3.1 Notes et comptes-rendus des séances du séminaire
D3.2 Recueil d’article pour publication chez MIT Press (collection Infrastructure) ou OE Press
D3.3 Rapport d’étape
D4 Rapport final

Données du projet

Un plan de gestion des données – notamment les réponses aux questionnaires, transcriptions d’entretiens, logs serveurs, contenant des données personnelles non sensibles – sera réalisé au lancement du projet, soumis pour validation au Service de protection des données du CNRS par la plateforme Revcil, déposé sur DMP OPIDOR, mis en ligne et mis à jour tous les 6 mois. Certains jeux de données existants pourront être réutilisés : données du projet CPJ (ex. Dumas Primbault, 2023b et à venir), données de l’Observatoire des usages d’OE Lab, données du projet ANR SO-DRIIHM (Dayre *et al.*, 2023), données de l’étude “Décliner la science ouverte” (Gruson-Daniel *et al.*, 2022b).

Personnes impliquées

Partenaire	Nom	Prénom	Position actuelle	Rôle dans le projet	Implication sur la durée du projet
OpenEdition	Dumas Primbault	Simon	CPJ	Coordinateur WP 1, 2, 3	15 p.mois
OpenEdition	Mounier	Pierre	IR	WP 1, 3	5 p.mois
OpenEdition	Guignonis	Sandra	IR	WP 2	5 p.mois
OpenEdition	1 postdoctorant PaRéDo à recruter			WP 1, 2, 3	24 p.mois
CIS	Galan	Jean-Marc	CRCN	Collaborateur WP 1, 3	10 p.mois
CIS	Musiani	Francesca	DR	WP 2, 3	5 p.mois
CIS	1 IR PaRéDo à recruter			WP 1, 2, 3	24 p.mois

Calendrier

		2024	2025				2026			2027	
		Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	
Étapes préliminaires	Recrutements	■									
	Kick-off	■									
	PGD		■		■		■		■	■	
WP1	T1.1		■	■							
	T1.2			■	■						
	T1.3				■	■					
	T1.4					■	■				
WP2	T2.1		■	■	■	■					
	T2.2			■	■	■	■				
	T2.3						■	■			
WP3	T3.1		■	■	■						
	T3.2					■	■	■			
Clôture du projet	Suivi livre					■	■	■			
	Rapport final								■	■	
	Clôture									■	

Budget – aide demandée à l'ANR

		OE Lab (gestion AMU)	CIS (gestion CNRS)
A. Frais de personnel		Jan. 2025 - déc. 2026 1 postdoctorant (< 2 ans d'expérience) 116.088 €	Jan. 2025 - déc. 2026 1 IE expérimenté (< 5 ans) 102.000 €
B. Coûts des instruments et du matériel (APCs, livre)		8.000€	0
C. Coût des bâtiments et des terrains		0	0
D. Coût du recours aux prestations de service		0	0
E. Frais généraux non forfaitisés	Conférence de lancement	7.000 €	0
	Séminaire GDKI	15.000 €	0
	Missions du personnel	3.000 €	3.000 €
	Conférence de clôture	0	7.000 €
Frais de gestion et de structure	0.145 * (A+B+C+D+E)	21.617,76 €	16.240 €
Sous-total		170.705,76 €	128.240 €
Total		298.945,76 €	

Impact et retombées du projet

Le WP1 du projet ambitionne de rassembler pour la première fois en un même corpus toutes les littératures – académiques, politiques et techniques – sur la question des données en SHS (D1.1). Ce paysage discursif dans des genres très différents permettra de **mieux comprendre la dynamique des prises de**

positions à l'égard des données ouvertes, des structurations de communautés autour de problématiques particulières, des formes de gouvernances, politiques et infrastructurelles, à la fois en tant qu'elles sont visées par ces critiques et en tant qu'elles tentent d'y répondre et d'équiper les pratiques au plus juste (D1.2, conférences et publications dans des revues à comité de lecture).

Au cœur du projet, le WP2 de "recherche sur la recherche" permettra tout d'abord la réalisation d'un **état des lieux panoramique – toutes disciplines, tous métiers confondus et tout au long du cycle de vie des matériaux – des pratiques de données ouvertes en SHS en 2025-2026** (D2.1, D2.3, conférences et publications dans des revues à comité de lecture). Les données produites dans ce cadre pourront être réutilisées pour répondre à d'autres problématiques, tandis que le protocole d'enquête sera modulaire, répliquable et ouvert et pourra ainsi être déployé sur d'autres terrains (D2.2). Encore jamais réalisé à cette échelle, cet état des lieux permettra en particulier de **documenter en détail les freins épistémologiques au partage et à la réutilisation de données** en SHS (D2.3).

Sur un versant opérationnel, le WP3 permettra d'émettre des **recommandations quant aux leviers techniques et aux leviers de gouvernance** qu'il sera possible de déployer afin de débloquent ces freins. Ces préconisations s'échelonnent sur un spectre allant de **préconisations très concrètes que COMMONS pourra mettre en place au cours de son développement** et de son déploiement, à des **propositions plus générales sous la forme de structures de gouvernance** adaptées à des matériaux et des pratiques ciblées, ainsi que des **suggestions pour adapter les principes de l'open data à la diversité épistémologique des SHS** (D3.2, D3.3, conférences et publications dans des revues à comité de lecture).

Plus largement, une étude de terrain d'une telle ampleur, liant les pratiques réelles des SHS aux effets produits sur celles-ci par les infrastructures permettra de **répondre à certaines critiques émises par la communauté académique à l'égard de certains principes de science ouverte et de l'open data**, en particulier sous l'angle de la datafication, et ainsi de **favoriser, grâce à leur implication, l'acculturation des communautés de recherche en SHS, autant qu'en STM, à l'exposition de leurs données**, au lien entre données et publications ainsi qu'au partage et à la réutilisation des données (D4).

Bibliographie

- Akers, Katherine G. et Jennifer Doty (2013). "Disciplinary Differences in Faculty Research Data Management Practices and Perspectives", *The International Journal of Digital Curation*, 8(2), 5-26, <https://ijdc.net/index.php/ijdc/article/view/8.2.5/332>
- Albornoz, Denise, Maggie Huang, Issra Marie Martin, Maria Mateus, Aicha Yasmine Touré et Leslie Chan (2018). "Framing Power: Tracing Key Discourses in Open Science Policies", *ELPUB 2018 Proceedings*, <https://dx.doi.org/10.4000/proceedings.elpub.2018.23>
- Baillot, Anne (2016). "A certification model for digital scholarly editions: Towards peer review-based data journals in the humanities", *Digital Scholarly Editing: Theory, Practice, Methods*, Université d'Anvers, Oct 2016, Anvers, Belgium, <https://shs.hal.science/halshs-01392880>
- Bardiot, Clarisse, Esther Dehoux et Emilien Ruiz (dir.) (2022). *La fabrique numérique des corpus en sciences humaines et sociales*, Lille, Presses Universitaires du Septentrion, http://www.septentrion.com/html/WYSIWYG/files/files/CONTENEUR2108_Bardiot_Dehoux_Ruiz.pdf
- Bendjaballah, Selma et Guillaume Garcia (dir.) (2023). "Les sciences sociales à l'épreuve de l'ouverture des données de la recherche", dossier spécial de *Terrains & Travaux*, 2(43), <https://doi.org/10.3917/tt.043.0211>
- Bonde Thylstrup, Nanna, Mikkel Flyverbom et Rasmus Helles (2019). "Datafied knowledge production: Introduction to the special theme", *Big Data & Society*, 6(2), <https://doi.org/10.1177/2053951719875985>
- Borgman, Christine (2020) [2016]. *Qu'est-ce que le travail scientifique des données ?*, Marseille, OpenEdition Press, <https://doi.org/10.4000/books.oep.14692>
- Cabrera, Francesca (2014). "Les données de la recherche en Sciences humaines et sociales : enjeux et pratiques. Enquête exploratoire", Mémoire de master en SIC, CNAM, https://memsic.ccsd.cnrs.fr/mem_01117375

- Dayre, Pascal, Émilie Lerigoleur, Bertrand Morandi (2023). “Données de l'enquête "Partagez votre expérience de partage de données !" (juin 2021) - projet ANR SO-DRIIHM”, NAKALA, <https://doi.org/10.34847/nkl.db933xfo>
- Dumas Primbault, Simon (2023a). “Naviguer dans les savoirs à l'ère numérique. Pour une ethnographie des pratiques informationnelles sur Gallica”, *Études de communication*, 61, 61-89, <https://doi.org/10.4000/edc.16108>
- Dumas Primbault, Simon (2023b). “Transcriptions de sept entretiens semi-directifs réalisés avec des usagers de Gallica au sujet de leurs pratiques informationnelles”, NAKALA, <https://doi.org/10.34847/nkl.320eu55q>
- Dumas Primbault, Simon (2024). “Documenter la navigation en bibliothèque numérique. Retour sur un chassé-croisé méthodologique entre qualitatif et quantitatif”, à paraître.
- Féret, Romain, Françoise Catherine Gouzi, Sandra Guignonis, Hélène Jouguet et Nicolas Larrousse (2021). “Guidelines for journals that wish to establish a “data policy””, Comité pour la science ouverte, <https://dx.doi.org/10.52949/19>
- Fisher, Angelina et Thomas Streinz (2022). “Confronting Data Inequality”, *Columbia Journal of Transnational Law*, 60(3), 829-956, <https://dx.doi.org/10.2139/ssrn.3825724>
- Fricker, Miranda (2007). *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford, Oxford University Press, <https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>
- Frischmann, Brett M., Michael J. Madison et Katherine J. Strandburg (dir.) (2014). *Governing Knowledge Commons*, Oxford, Oxford University Press, <https://doi.org/10.1093/acprof:oso/9780199972036.001.0001>
- Gitelman, Lisa (dir.) (2013). “Raw Data” is an Oxymoron, Cambridge MA, MIT Press.
- Goëta, Samuel (2024). *Les données de la démocratie. Open data, pouvoirs et contre-pouvoirs*, Caen, C&F Editions.
- Gruson-Daniel, Célya et Groupe Projet Réussir L'Appropriation De La Science Ouverte (2022a). “Décliner la science ouverte : Rapport final”, Comité pour la science ouverte, <https://dx.doi.org/10.52949/27>
- Gruson-Daniel, Célya, Claire Lemercier, Anne Vanet, Juliette Hueber, Hélène Chambefort et Marie Herbet Marie (2022b). “Étude “décliner la science ouverte” : données issues du questionnaire “données et science ouverte” (2021)”, Recherche Data Gouv, V1, <https://doi.org/10.57745/V64RYT>
- Guignonis, Sandra, Nicolas Larrousse et Dominique Roux (2023). “Lier des données et des publications dans COMMONS. Le cas des *data papers*. Pratiques et usages”, Journée d'étude MSH Lorraine, <https://msh-lorraine.fr/nos-services/data-papers/>
- Hess, Charlotte et Elinor Ostrom (dir.) (2007). *Understanding Knowledge as a Commons: From Theory to Practice*, Cambridge, MA, MIT Press, <https://doi.org/10.7551/mitpress/6980.001.0001>
- Horsley, Nicola et Mike Priddy (2018). “D3.1 Report on Historical Data as Sources”, K-PLEX Project report, https://kplexproject.files.wordpress.com/2018/06/kplex_deliverable-d3-1.pdf
- Kaabachi, Bayrem et Simon Dumas Primbault (2023). “A Topological Data Analysis of Navigation Paths within Digital Libraries”, *CEUR Workshop Proceedings*, Computational Humanities Research, 3558, 111-134, <https://hal.science/hal-04301567>
- Kitchin, Rob (2022). *The Data Revolution: A Critical Approach to Big Data, Open Data, and Data Infrastructures*. Londres, Sage, <https://doi.org/10.4135/9781473909472>
- Leonelli, Sabina (2015). “What Counts as Scientific Data? A Relational Framework”, *Philos Sci*, 82(5), <https://doi.org/10.1086%2F684083>
- Leonelli, Sabina (2018). “Re-Thinking Reproducibility as a Criterion for Research Quality”, *Research in the History of Economic Thought and Methodology*, 36(B), 129-146, <http://philsci-archive.pitt.edu/14352/>
- Leonelli, Sabina (2022). “Open Science and Epistemic Diversity: Friends or Foes?”, *Philosophy of Science*, 89(5), 991-1001, <https://doi.org/10.1017/psa.2022.45>
- Leonelli, Sabina (2023). *Philosophy of Open Science*, Cambridge, Cambridge University Press, <https://doi.org/10.1017/9781009416368>
- Leonelli, Sabina et Nicolò Tempini (dir.) (2020). *Data Journeys in the Sciences*, Cham, Springer, <https://doi.org/10.1007/978-3-030-37177-7>
- Lund, Rebecca, Jill Blackmore et Julie Rowlands (2022). “Epistemic governance of diverse research practices and knowledge production: an introduction”, *Critical Studies in Education*, 63(5), 535-548 <https://doi.org/10.1080/17508487.2022.2136226>

- Mabi, Clément, Jean-Christophe Plantin et Laurence Monnoyer-Smith (dir.) (2017). *Ouvrir, partager, réutiliser. Regards critiques sur les données numériques*, Paris, Éditions de la Maison des sciences de l'homme, <http://books.openedition.org/editionsms/9026>
- Mohr, Alicia Hofelich, Josh Bishoff, Carolyn Bishoff, Steven Braun, Christine Storino et Lisa R. Johnston (2016). "When Data Is a Dirty Word: A Survey to Understand Data Management Needs Across Diverse Research Disciplines", *Bulletin of the Association for Information Science and Technology*, 42(1), 51-53, <https://doi.org/10.1002/bul2.2015.1720420114>
- Mostern, Ruth et Mariëka Arksey (2016). "Don't Just Build It, They Probably Won't Come: Data Sharing and the Social Life of Data in the Historical Quantitative Social Sciences", *International Journal of Humanities and Arts Computing*, 10(2), 205-224, <https://doi.org/10.3366/ijhac.2016.0170>
- Mounier, Pierre (2015). "Une « utopie politique » pour les humanités numériques ?", *Socio*, 4, <https://doi.org/10.4000/socio.1451>
- Mounier, Pierre (2018). *Les humanités numériques : Une histoire critique*, Paris, Éditions de la Maison des sciences de l'homme, <https://doi.org/10.4000/books.editionsms.12006>
- Mounier, Pierre et Simon Dumas Primbault (2023a). "Sustaining Knowledge and Governing its Infrastructure in the Digital Age: An Integrated View", pre-print, <https://zenodo.org/doi/10.5281/zenodo.10036401>
- Mounier, Pierre et Simon Dumas Primbault (2023b). "Governing Digital Infrastructures: Exploring a Key Issue at the Crossroads of Digital Humanities and Open Science", littérature grise de séminaire, <https://zenodo.org/doi/10.5281/zenodo.10036336>
- Mounier, Pierre et Simon Dumas Primbault (2023c). "Sustaining Knowledge and its Infrastructure: Digital knowledge infrastructures at the crossroads of governance practices", littérature grise de workshop, <https://zenodo.org/doi/10.5281/zenodo.10036367>
- Musselin, Christine (2021). "University Governance in Meso and Macro Perspectives", *Annual Review of Sociology*, 47, 305-325, <https://doi.org/10.1146/annurev-soc-090320-012708>
- Plantin, Jean-Christophe et Andrea Thomer (2023). "Platforms, programmability, and precarity: The platformization of research repositories in academic libraries", *New Media & Society*, <https://doi.org/10.1177/14614448231176758>
- Prost, Hélène et Joachim Schopfel (2016). "Les données de la recherche en SHS. Une enquête à l'Université de Lille 3", Rapport de recherche, Université Lille 3, <https://hal.univ-lille.fr/hal-01198379>
- Serres, Alexandre, Marie-Laure Malingre, Morgane Mignon, Cécile Pierre et Didier Collet (2017). "Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2", Rapport de recherche, Université Rennes 2, <https://hal.science/hal-01635186>
- Strasser, Bruno (2012). "Data-driven sciences: From wonder cabinets to electronic databases", *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(1), 85-87, <https://doi.org/10.1016/j.shpsc.2011.10.009>
- Thoergensen, Jennifer L. (2018). "'Yeah, I Guess That's Data': Data Practices and Conceptions among Humanities Faculty", *portal: Libraries and the Academy*, 18(3), 491-504, <https://eric.ed.gov/?id=EJ1185990>
- Tóth-Czifra, Erzsébet (2020). "The Risk of Losing the *Thick Description*: Data Management Challenges Faced by the Arts and Humanities in the Evolving FAIR Data Ecosystem", in Jennifer Edmond (dir.), *Digital Technology and the Practices of Humanities Research*, Cambridge, OpenBook Publisher, <https://doi.org/10.11647/OBP.0192>
- Vadrot, Alice B. M. (2011). "Conceptual framework for "Epistemic Governance"", *International Journal for Social Ecology and Sustainable Development*, 1(2), 44-53, <https://doi.org/10.4018/jesed.2011010104>
- Visvanathan, Shiv (1997). *A Carnival for Science: Essays on Science, Technology and Development*, New Delhi, Oxford University Press.
- Wiggins, Chris et Matthew L. Jones (dir.) (2023). *How Data Happened. A History from the Age of Reason to the Age of Algorithms*, New York, Norton and Company.