



**HAL**  
open science

## **Edito : Naissance d'une revue dédiée aux données en SHS**

Florence Thiault, Audrey Knauf

► **To cite this version:**

Florence Thiault, Audrey Knauf. Edito : Naissance d'une revue dédiée aux données en SHS. 2025, Les articles de données en., <10.46298/dc.17044>. <hal-05398681v2>

**HAL Id: hal-05398681**

**<https://hal.science/hal-05398681v2>**

Submitted on 11 Dec 2025

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 CC BY 4.0 - Attribution - International License



---

Éditorial

## Naissance d'une revue dédiée aux données en SHS

---

**Florence Thiault**, rédactrice en chef de *Data & Corpus – La revue des données en SHS*

**Audrey Knauf**, rédactrice en chef adjointe de *Data & Corpus – La revue des données en SHS*

**Date de publication** : 11 décembre 2025

© The Authors

Licence : [Creative Commons Attribution International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Publié par l'Université de Lorraine

**DOI** : <https://doi.org/10.46298/dc.17044>

Nous avons le plaisir de vous présenter le premier numéro de *Data & Corpus – La revue des données en SHS*. Cette revue naît d'un double constat : si les données sont au cœur de la recherche aujourd'hui, elles demeurent encore trop souvent invisibles dans les circuits traditionnels de publication ; et dans les sciences humaines et sociales (SHS), leur valorisation se heurte à plusieurs défis spécifiques, tels que ceux liés à la diversité disciplinaire, au caractère qualitatif et sensible des corpus, ou encore aux contraintes éthiques, juridiques et techniques.

C'est pour répondre à ce besoin qu'un collectif interdisciplinaire et national de chercheur·es, d'ingénieur·es, de documentalistes et d'éditeur·rices a initié en 2022 le projet *Data & Corpus*. Son objectif est d'offrir un espace éditorial entièrement dédié aux articles de données (*data papers*) en SHS, en accès ouvert diamant, c'est-à-dire diffusés sans frais de publication et accessibles à tous·tes. La revue se donne ainsi trois ambitions. Elle entend d'abord soutenir la science ouverte en SHS en rendant visibles et réutilisables les données de recherche dans le respect des principes FAIR et des exigences éthiques propres à nos disciplines. Elle vise également à promouvoir l'interdisciplinarité en décloisonnant les approches et en favorisant les échanges autour des méthodes de production et de documentation des données. Enfin, elle affirme la nécessité de reconnaître le travail collectif en valorisant l'apport conjoint des chercheur·es, ingénieur·es, archivistes, documentalistes et autres professionnel·les qui contribuent à la collecte, à la structuration et à la mise en circulation des données.

Le lancement officiel de la revue s'est accompagné, le 31 mars 2025, d'une [journée d'étude](#) sur « *Les articles de données en sciences humaines et sociales : enjeux sociétaux et innovations méthodologiques* » organisée à la MSH Lorraine. Cette rencontre scientifique, qui faisait suite à la journée de mars 2023 consacrée à la genèse d'un data journal interdisciplinaire en SHS, avait pour objectif de présenter les projets d'articles qui composent ce premier numéro.

Les communications ont mis en lumière la richesse et la diversité des données et des corpus en SHS, mais aussi les enjeux méthodologiques et sociétaux qui leur sont associés.

Les six articles rassemblés ici illustrent ces ambitions.

- Guillaume Levrier et Dorothee Benhamou-Suesser présentent *les campagnes disparues*, un travail pionnier sur les archives électorales collectées par la BnF depuis les années 2000. Grâce à l'indexation plein texte et au logiciel libre PANDORÆ, ces données deviennent enfin accessibles aux chercheurs, ouvrant la voie à des analyses inédites des débats publics en ligne.

- Fanny Arnaud et ses collègues proposent une base unique de *photographies aériennes obliques* documentant le Rhône entre 1919 et 1961. Ce corpus, enrichi de métadonnées et diffusé en photothèque interactive, offre un outil précieux pour comprendre les trajectoires socio-environnementales du fleuve et constitue un support de valorisation auprès des acteurs locaux et du grand public.
- Aurélie Nardy, Isabelle Rousset et Rachel Gaubil s'intéressent au *recueil non supervisé de données audio massives en école maternelle* dans le cadre du projet ANR DyLNet. En combinant enregistrements, transcriptions et développement d'outils logiciels, elles rendent disponibles un corpus inédit de 705 transcriptions permettant de mieux comprendre la socialisation langagière des enfants entre 3 et 6 ans.
- Maxime Crépel et ses co-auteur·es présentent une *cartographie web de l'écosystème médiatique français*. Leur base de données, rassemblant 747 sites de médias enrichis de métadonnées variées, constitue une ressource de premier plan pour analyser la diversité médiatique, ses transformations et ses enjeux démocratiques.
- Marie Skrovec, Flora Badin et Céline Dugua mettent en lumière le passage d'*ESLO* à *ESLO-FLEU*, en valorisant le patrimoine linguistique du corpus orléanais pour l'enseignement du français langue étrangère. Leur sélection de 150 extraits sonores thématiques et annotés ouvre des perspectives originales à la fois pour la recherche sociolinguistique et la didactique du FLE.
- Victor Gay propose un retour critique sur la *base de données Jean Nicolas*, recensant plus de 8 500 rébellions survenues en France entre 1661 et 1789. En évaluant les biais méthodologiques de cette ressource et en proposant des pistes d'usage raisonné, il contribue à renforcer sa valeur pour les historien·nes et les sciences sociales.

Si le *data paper* est encore relativement peu répandu en SHS, les contributions rassemblées dans ce numéro montrent qu'il contribue au travail scientifique en renforçant la documentation et la traçabilité des données. Qu'il s'agisse de rendre explorables les archives électorales du web, d'ouvrir l'accès à des corpus volumineux et sensibles en milieu scolaire, de cartographier les médias contemporains, de documenter l'évolution d'un fleuve, de transmettre un patrimoine linguistique ou de revisiter les archives d'un historien, chacun de ces travaux illustre la capacité des données à identifier de nouvelles pistes de recherche et à enrichir les *communs* qu'ils soient scientifiques, numériques ou de la connaissance.

La revue *Data & Corpus* souhaite également valoriser des travaux liés aux *data papers* en SHS. Pour ce premier numéro, nous vous proposons le

compte rendu, rédigé par Pierre Peraldi-Mittelette, de l'ouvrage « Publier, partager et réutiliser les données de la recherche : les *data papers* et leurs enjeux », dirigé par Christine Kosmopoulos et Joachim Schöpfel aux Presses universitaires du Septentrion (2025). Composé de 18 chapitres, ce livre constitue une référence majeure pour comprendre les *data papers* en SHS.

Nous espérons que ce numéro inaugural saura nourrir les débats, inspirer de nouvelles initiatives et encourager l'appropriation de ce format éditorial émergent. Plus largement, nous souhaitons qu'il contribue à faire reconnaître pleinement le rôle des données comme objets scientifiques à part entière, à rendre visible le travail essentiel qu'implique leur traitement, et qu'il accompagne la communauté SHS vers des pratiques de recherche transparentes, collaboratives et pérennes.