



**HAL**  
open science

# Navigating Knowledge in the Digital Age: A Study of Researchers' Information-Browsing Practices in Digital Libraries

Simon Dumas Primbault

► **To cite this version:**

Simon Dumas Primbault. Navigating Knowledge in the Digital Age: A Study of Researchers' Information-Browsing Practices in Digital Libraries. 2022. hal-03727166

**HAL Id: hal-03727166**

**<https://hal.science/hal-03727166>**

Preprint submitted on 19 Jul 2022

**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.

# Navigating Knowledge in the Digital Age

A Study of Researchers' Information-Browsing Practices in Digital Libraries

Simon Dumas Primbault

École polytechnique fédérale de Lausanne

Laboratory for the History of Science and Technology (LHST)

[simon.dumasprimbault@epfl.ch](mailto:simon.dumasprimbault@epfl.ch)

## Table of Contents

Summary	2
Proposed research	3
State of the fields	3
History and anthropology of knowledge	3
Digital humanities and data sciences	3
Information and communication studies	4
On navigation	5
On arXiv and Gallica	5
State of personal research	5
Detailed research plan	6
A “methodological criss-crossing”	6
Research Objective 1: Cross-Histories of two Digital Libraries	8
Research Objective 2: Digital Libraries as Technical Media	9
Research Objective 3: An Ethnography of Digital Navigation	11
Schedule and milestones	14
Relevance and impact	15
Bibliography	16

## Summary

The birth and development of digital libraries—broadly understood as curated collections of electronic documents accessible online on dedicated platforms with tools for search and consultation—have radically transformed research activities. From any computer with Internet access, the wealth of information available allows for the simultaneous consultation of a great variety of resources and for their continual rearrangement into renewed information landscapes. Consequently, traditional practices observed in physical places of knowledge such as libraries and archives—searching catalogues, browsing through shelving, taking notes—have been supplemented with a series of digital practices developed by researchers to browse websites and databases—searching by keywords, filtering results, navigating through links.

While historically science was, and to a great extent still is, arranged into arborescent taxonomies, scholars have in practice always negotiated with this normative order of knowledge, resulting in a complex landscape much thicker than a tree-like structure. In contributing to the digitization of information practices, digital libraries are further redrawing this landscape and its relation to established orders of knowledge. The appropriation of digital libraries by their users opens the “tree of knowledge” to a multiplicity of other dimensions allowing for leaps from one “leaf” to another, as well as a continuous reconfiguration of branches of knowledge. Once we become aware that the *practical* landscape of knowledge is evolving into a much more dynamic arrangement, the use of spatial concepts to understand research practices *from the point of view of scholars* becomes essential.

Going beyond the blunt dismissal of cyberspace as a collective illusion, the ambition of this project is to reclaim a spatial understanding of digital libraries through the investigation of *navigation* as a fundamental information practice for researchers. This project proposes a critical reappraisal of the digital as a space ceaselessly generated by the very navigational practices of users. The aim is to recover the three dimensions of digital libraries as political spaces—with their governance resulting in choices and hierarchization of content—, mediatic spaces—made of layers of mediation framing their structure and use—, and experiential spaces—appropriated by researchers in practice. By documenting digital knowledge spaces in-the-making through the prism of navigation, the goal of this project is to address what it means to “orient oneself in thinking” in the digital age.

To that end, I will look at two digital libraries as case studies accounting for two different kinds of engagement with electronic documents: **arXiv**—an e-print repository for natural sciences hosted by Cornell University—and **Gallica**—the online platform for the collections of the Bibliothèque nationale de France. While arXiv is a born-digital medium for the circulation of electronic scientific papers among a scholarly community involved in its maintenance and development, Gallica was construed as the online platform of a public patrimonial institution, tailored to make available digitized historical sources to the widest audience. Such a comparative approach between a digital library (of archival material) and a repository (fashioned as an archive) will help me understand (i) the information specificities of sciences, social sciences, and humanities; (ii) the putative convergence as data of primary sources and secondary literature; (iii) public institution or self-regulated community as different modes of governance; (iv) how open access is negotiated and implemented in different cultural, political, and technological contexts.

This project proposes a novel methodological criss-crossing unfolding in three times, each corresponding to a different disciplinary perspective and addressing one of the dimensions—political, mediatic, and experiential—of digital libraries. This innovative mixed-methodological approach will not only allow to account in an integrated fashion for the thickness of digital navigation seen from different points of view, but it will also permit the mutual critique of perspectives, each highlighting the assumptions and blind spots of the others. **Research Objective 1 will be dedicated to studying the cross-histories of arXiv and Gallica from their creation in the 1990s until today**, with a vast array of source materials: administrative archives, oral testimonies, and Internet archives. **Research Objective 2 will aim to conduct a media study of both digital libraries**, descending through their layers of mediation from interface through information architecture to code. **Research Objective 3 will consist in carrying an in-depth ethnography of researchers’ navigational practices** by combining qualitative methods—semi-directed interviews and non-participant observations—with a quantitative analysis of server logs—using topological data analysis to study the morphology of navigational paths. Partaking in the understanding of digital knowledge infrastructures, this research could be involved in their design by providing tools to analyse navigation or by advising policies.

# Proposed research

## State of the fields

This research project comes within the scope of the following fields, which are neither mutually exclusive, nor impervious to each other as would suggest the following partition made only for the sake of clarity. Among the bounteous literature, emphasis was put on practices and technologies of information management, and more specifically on navigation.

### *History and anthropology of knowledge*

The history of information management developed in the course of the last decades parallel to the re-emergence of concerns about information overload due to the ever more pervasive presence of digital media in daily life. And in scholarly activity as well, since it prompted historians of knowledge to gaze back at the past from the perspective of the changes brought about in libraries by the irruption of “the machine” (Tresch, 2014). Initially rooted in the anthropology of writing and the material “technologies of intellect”, such as tables, lists, or formulae as studied by Jack Goody (1977), the history of information management first endeavoured to document the writing practices and graphical devices used through time to fix, accumulate, and circulate knowledge. The concept of inscriptions was later reappraised by Bruno Latour in his laboratory studies, thereby defining an anthropology of knowledge focused on the combination and cascade of “immutable mobiles” used in the social construction of facts through conviction (Latour, 1986).

More recently, media studies have appealed to historians and anthropologists of knowledge, who emphasized the materiality of communication and the process of stabilising a natural referent through the construction of a network of inscriptions, which the scholar navigates (Lenoir, 1997). This led to the perspective of an archaeology, borrowed from Foucault and reappraised by Kittler (1990), grounded in three main aspects (Citton, 2017). Firstly, the archaeology of a technology of intellect is concerned first and foremost with the material aspect of the technology—rather than taking as its primary object the discourses held by and produced on it—, and with the bodily and cognitive effects of this materiality on scholarly life. Secondly, and as a direct result of this instrumental approach, an archaeology proposes to map the scholarly practices both permitted and constrained by the materiality of their tools rather than aiming at the construction of a narrative. Finally, an archaeology considers its object from a synchronic perspective, rather, it seeks the diachrony within the synchronicity through the stratigraphy of the successive layers of practices and materialities coexisting in its object of study—paper and screen, skimming and thorough reading, targeted searches and exploratory browsing.

The long cultural history of information technologies, from the Middle-Ages to the contemporary era through Renaissance and the industrial period, has since then become an established field (Blair *et al.*, 2021). Coupled with a material history of knowledge practices (Bert and Lamy, 2021), they shed light on the tight entanglement, through times, of practices and media, of information systems and behaviours, of users and technologies. Transposing these issues in the late 20<sup>th</sup> and early 21<sup>st</sup> centuries entails to weave this media-historical gaze (Gitelman, 2014; Krajewski, 2018) with other disciplines.

### *Digital humanities and data sciences*

Research on digital scholarly practices in the 90s and the early 00s structured around “information-retrieval behaviour” (see Case (2002) for a detailed synthesis). Mostly ethnographic in method, this research focused on early “electronic journals”, rather than fully-fledged digital libraries as whole infrastructures, and took as its unit of analysis the Web page, rather than the document with its metadata and the operations it may be subjected to.

The digital humanities (DH) and data science entered the scene in the late 2000s focusing more specifically on “interaction traces” (understood as the sequential records of a user’s interactions with one or across multiple platforms). The birth and development of the Semantic Web led to a lot of research on the part of data and computer scientists in an effort to identify, understand, and manage digital information-seeking as well as the traces it leaves behind—Wikipedia being the most extensively covered platform by the existing literature (e.g., Piccardi *et al.*, 2020). Aiming at collecting, quantifying, and modelling online traces, the investigation of interaction traces helped emphasize the role of navigation, as opposed to search, in seeking

information: on the one hand, “the perfect search engine is not enough” therefore prompting user to browse (Teevan, 2004), while, on the other hand, existing recommendation algorithms do not favour navigability, even hindering it (Lamprecht, 2015).

Yet, inheritors of a rather mechanical model supported by “information foraging theory”, these studies are still too often dedicated to practical optimisation and problem-solving—aimed at enhancing the ranking of pages found through “post-query navigation” (Bilenko and White, 2008; Singla *et al.*, 2010), predicting the next page based on statistical regularities (Sen and Hansen, 2003; Koopman *et al.*, 2019), or suggesting the most beaten “trails” (White and Huang, 2010). Furthermore, when users are brought to the fore by data science and UX studies, the emphasis is usually put on “directed search”—*i.e.*, shortest, directed, and local paths—, thereby neglecting a whole part of navigation strategies: crawling, exploratory searches, serendipity...

When eschewing this way more traditional ethnographic approaches based on interviews and on-site observations, navigation is deemed worthy of practice and study only for its end products or its “waypoints” (White and Singla, 2011), not as a process in itself, and it is assessed according to relevance only, rather than discovery or originality. Most quantitative studies on navigation therefore bypass altogether a practice that their authors nonetheless underlined as fundamental: search engines, however accurate and powerful, never fully satisfy users who tend to rely more on step-by-step contextual navigation.

Within the realm of DH, the recent development of topological data analysis (TDA) calls for a reappraisal of navigation through the use and honing of tools specifically devoted to the qualitative study of shape. The shape of the World Wide Web is an issue as old as the Internet itself and still widely debated to this day (Ghitalla, 2021). Traditionally, the Web has been modelled as a graph. Although this has proven to be a powerful tool to study digital communication, graphs are intrinsically limited to model pairwise interactions. The recent success of topological methods in studying data, and the parallel establishment of topological data analysis as a field (Edelsbrunner and Harer, 2010), have confirmed the utility of viewing data through a higher-dimensional analogue of graphs. Providing the tools to model navigation as a rich and thick process,<sup>1</sup> rather than as a problem to solve, TDA offers the possibility to understand it in a less systematic, more descriptive way, if coupled with humanistic methods addressing navigation in its lived practical thickness. To our knowledge, although TDA has been used to analyse scientific collaborations (Patania *et al.*, 2017), nothing properly topological has been endeavoured for navigational practices yet.

### *Information and communication studies*

While the anthropology of knowledge has long addressed material practices in the making of science in physical libraries, laboratories, and archives (e.g., Daston, 2017), library ethnographies of the 2000s and early 2010s have also proved to be more “ethnographish” satisfaction surveys oriented towards problem-solving (Lanclos and Asher, 2016)—with a recent emphasis on user experience, or UX (Priestner and Borg, 2016)—, rather than open-ended long-term research.

However, since their definition in the late 90s, notably by Christine Borgman (2000), *digital* libraries have become the subject of many international conferences (e.g., TPDL or ICCEDL) and dedicated journals or platforms (e.g., *portal* or the *International Journal on Digital Libraries*). They have also become a privileged object of inquiry for information and communication studies. As a result, since the late 2010s, within the field of information studies, a growing body of interdisciplinary research at the crossroads of ethnography, media studies, history, and digital humanities has focused on scholarly practices at large in the digital era. More specifically interested in the practices of digital humanists (Antonijević, 2015), and of humanists adopting the computer (Trace and Karadkar, 2017; Given and Wilson, 2018), these studies aim at mixing empirical methods towards an ever more integrated understanding of scholarly work in its extreme diversity of practices and medias. From the psychologizing vocabulary (needs, behaviour...) of late twentieth-century studies on information-retrieval, the focus has shifted to early 2020s “scholarly workflow” (Antonijević, 2020; Weiland, 2021) of researchers become “bricoleurs” (Antonijević and Cahoy, 2018). So much so that young scholars may now learn in handbooks that their practice is “a nonlinear body of research—some of it orderly and rational, much of it chaotic and contingent, all of it loosely tied together” (Abbott, 2014: 35).

---

<sup>1</sup> In the same manner T.D. Anderson integrated “relevance as process” (2005) and “uncertainty as action” (2006) within scholarly practices.

### *On navigation*

The present project follows this trend, focusing on one of the many roles of the scholar as researcher, through the concept of navigation encompassing multiple practices and medias. As early as 1989, Marcia J. Bates proposed “berrypicking” as a counterpart to information-retrieval (Bates, 1989). Yet, despite this seminal article and scarce interest in the 2000s,<sup>2</sup> browsing or navigation, apart from being named in long lists of research practices, have received little attention from information studies.

After an early fame with the birth of the Internet, spatial metaphors have later been handled cautiously, due to their excessive practicality being suspected of denoting an illusory territory—the cyberspace—that was being called into question. However, as long as we do not construe digital spaces as we do our physical environment, but instead we pay attention to how they are constructed by practices and their representations (Mounier, 2010), navigation becomes a relevant concept as recent interest shows. Indeed, reading and writing practices have evolved a lot over the past decades, with some authors reviving the picking model (Ghitalla *et al.*, 2003: 177-178) or the figure of the “hunter-gatherer” (Le Marec and Mairesse, 2017). Deep log analysis of scholarly publishers led by the CIBER Research group in the UK described reading as either “bouncing, flicking, or skittering” (Nicholas and Clark, 2012: 95), while the same year a mixed-methods study led by Chérifa Boukacem-Zeghmouri on STM French scholars showed the “fundamental nature of navigation for researchers” (2012: 89). More recently, a broad and international survey addressing information practices in all disciplines, highlighted the overwhelming prevalence of browsing through single websites and chaining references over keyword searching (Tenopir *et al.*, 2019). Endeavouring to document in depth the navigation of scholars within physical environments, McKay *et al.* (2019) conducted a mixed-methods inquiry that led to a typology of shelf browsing in libraries and policies on how to design digital ones. Yet, while a general interest in the topology of culture (Lury *et al.*, 2012; Boullier and Lévy, 2016), and more specifically of digital reading practices (Piper, 2013), has recently grown, nothing of such substance has been endeavoured regarding digital libraries.

### *On arXiv and Gallica*

Although a number of surveys are regularly conducted by both physical and digital libraries themselves, their results are seldom made public. There exist only a few studies on Gallica, among them one reflection on “digitized patrimony” in a book chapter aptly titled “to orient oneself among sources” (Roustan, 2016). More importantly, one in-depth mixed-method study on information practices in Gallica was designed (Beaudouin and Denis, 2014) and conducted until 2016 (see below). Gallica’s history is scarcely written.

Besides bits of its early history (McKiernan, 2000), several written by its founder Paul Ginsparg (2011, 2017, 2021), there is little research on arXiv, usually focused on the relationship between pre-prints and academic journals (e.g., Larivière *et al.*, 2014). Although a master’s thesis showed the interest of “using access data for paper recommendations on arXiv.org” (Pohl, 2006), it merely focused on matters of relevance and did not inquire about navigation as such.

## State of personal research

Rooted in the historical anthropology of knowledge, my doctoral research focused on the information practices of two seventeenth-century scholars. By delving into the personal archives—working papers, drafts, reading notes, sketches, preparatory manuscripts—of Gottfried Wilhelm Leibniz and Vincenzo Viviani (Galileo’s last disciple), I set out to document what the materiality of a specific medium (in this case ink and paper), as well as the various uses of said materiality, did to the process of thinking. Within a “written culture” largely centred on the book and the nascent article, scholarly reflection was deployed through drawing or mathematical symbolism on the two-dimensional space opened by the surface of paper, and research strategies to orient oneself in thinking relied on new paper tools (e.g., dictionaries, thesauri, commonplace books, collections).<sup>3</sup>

In order to document the production of knowledge in the early modern era, through the prism of media, I made an indirect use of historical sources to detect, through their materiality, the practices of notetaking, cutting and pasting, the

---

<sup>2</sup> Notably Bates (2007), albeit grounded in behavioural psychology, and Ghitalla (2002) whose rich definition of navigation as an activity coupled with a technical medium bridges the gap between “subjective geographies of information” and the co-construction of “common territories”.

<sup>3</sup> For the past two years, I have been teaching this as a lecture in the digital humanities on the long history of “information management”.

constitution of a library, etc. Rather than documenting writing practices *per se*, I studied how reading-writing technologies were used by scholars to organize the surface of their papers, sort their own archive, navigate their peers' work, and more generally orient themselves in the knowledge landscape of their time. These technologies of the intellect were very generally based on a taxonomic order of knowledge, *i.e.* trees of knowledge organised into disciplines and sub-disciplines, nested according to their objects or methods (Salonius and Bintley, 2021); classifications whose content has often been debated and rehashed, while the tree structure has remained the same until today.

This research led to the publication of peer-reviewed articles on Viviani's practice of drawing, and on how, after Galileo, his use of mathematical tools to describe physical objects contravened to the tree-like order of knowledge, thereby making his practice of mixed-mathematics a subaltern science. Furthermore, I studied how Viviani organized his work with lists, headings, and cross-references enabling him to "navigate" his own archive and his personal library conceived of as an "ink-and-paper milieu" (Dumas Primbault, 2020a, 2020c).<sup>4</sup> My work on Leibniz allowed me to contribute to the archaeology of computational reasoning by linking his material "form of thought" with the birth and development of an analytical-combinatorial regime of knowledge. My research showed how Leibniz's material practice of moving and combining "bits and pieces" of inscriptions on the surface of paper complemented his piecewise conception of "information" (Dumas Primbault, 2020b, 2021, 2022).

In 2019, transposing my doctoral research to the contemporary era, I earned together with J. Baudry and J.-F. Bert an interdisciplinary grant between EPFL and Unil on the topic of "digital mobilities". Entitled "Digital Trail-Blazing: Understanding and Remaking Intellectual Mobility on Online Research Platforms", the project was led by the Laboratory for the History of Science and Technology. In the footsteps of Beaudouin *et al.* (2016) and Nouvellet *et al.* (2017), this research situated at the crossroads of ethnography, data science, and digital humanities endeavoured to shed light on the navigational practices of researchers on Gallica. Bridging the gap between the *ethnography of the digital*—the qualitative study of scholars' practices through observations and interviews—and *digital ethnography*—the quantitative analysis of navigation traces—, it developed mixed methods combining interviews with topological analysis of navigation paths extracted from server logs. This exploratory study, whose code is on the laboratory's Github under a GNU license,<sup>5</sup> led to promising findings on researcher's navigation within the Dewey classification, on the role of "pivotal literature", and on a first sketch of a variety of "regimes of navigation" clustered based on their topological features (Dumas Primbault *et al.*, 2021b). These findings are the empirical foundations of the present project.

Finally, for more than two years I have been working with a team of developers, editors, and researchers under the supervision of Christian Jacob, on the development of the *Savoirs* platform. Founded on the principle of navigation, this platform is intended to be a digital library whose thematic, chronological, and geographical reconfiguration makes it possible to complicate the order of knowledge and to propose cross-readings, comparative approaches, unexpected associations of ideas, and original paths.<sup>6</sup> I have more specifically been in charge of the design and supervision of the development of recommendation algorithms based on the database's metadata and the current state of the reader's parameters.

## Detailed research plan

### *A "methodological crisis-crossing"*

Among the many roles of the scholar—alternately author, reviewer, advisor, editor, teacher, or administrator—their activity as researcher suffers no definitive model or framework. It rather appears to be a complex and undetermined skein woven from a multiplicity of integrated practices and medias, phases and objectives, regimes and strategies. The present project proposes to document the most recent history of information management following one thread of this tapestry: navigational practices. Although traditionally associated with hermeneutic disciplines like the humanities, the ever more piecewise access to information brought about by digital media has gradually made navigation a transversal practice. As a material practice conducted through

---

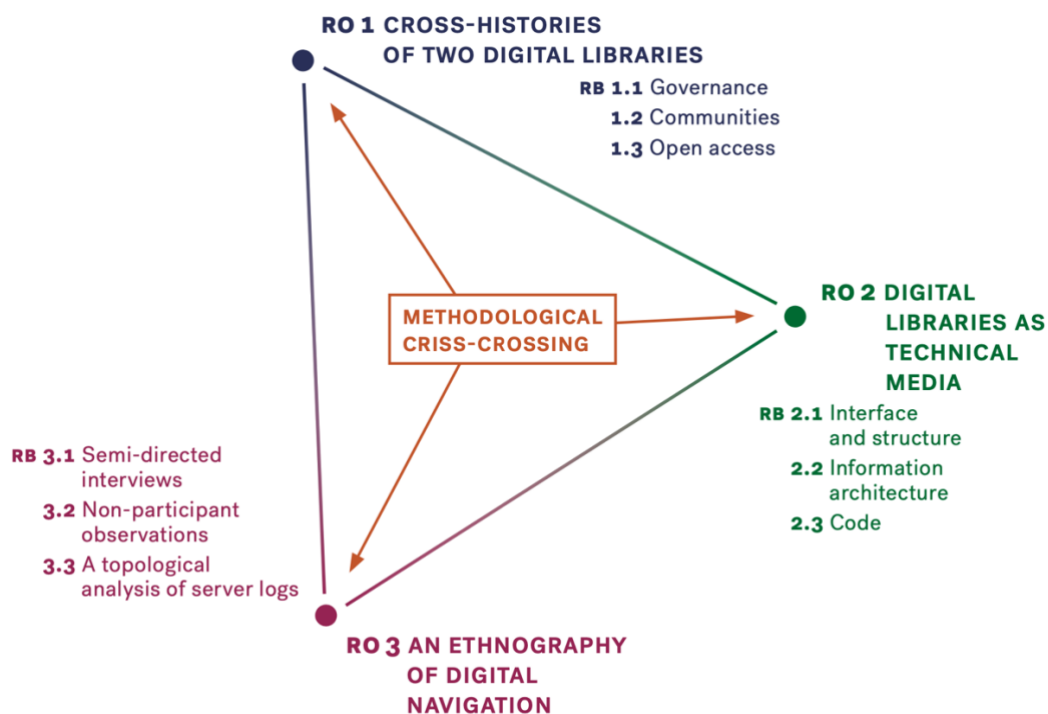
<sup>4</sup> My monograph on Viviani's activity as an engineer for the Duchy of Tuscany was submitted to the Éditions de l'EHESS and is currently under revision.

<sup>5</sup> <https://github.com/LHST-EPFL/TDA-Gallica>

<sup>6</sup> See <https://savoirs.app> and <https://savoirs.info> for more information about this project supported by EHESS, EPFL, ENSSIB, CNRS.

libraries, navigation does not amount to the unconstrained roaming of infinite spaces but must rather be understood as a patient work of contextual surveying within the framework imposed by their media. As such, it is a diachronic, albeit not simply sequential, practice aimed at finding contextual relationships between documents based on their types, topics, authors, disciplines, or methods. Navigation therefore encompasses a wide variety of information practices spanning searching (or information-mining), collecting (information-farming), skimming (information-strolling), garnering (information-stumbling, serendipity, or accident), and forays into neighboring fields (information-poaching, or “*braconnage*”). Thus, navigation is an exercise in orientation within knowledge: moving from one document to another, researchers learn to draw for themselves an information landscape by grouping works encountered on their path into disciplines, periods, or approaches. This activity implies reading (both the texts of documents and the “*architexte*” of the interface) and writing (both deliberate and unintentional traces of interaction left on server logs) and does not preclude other practices as it prompts notetaking, thorough writing, editing, and close reading.

The practice of navigation therefore appears as a multifaceted object and calls for a diversity of perspectives and methods of inquiry. To make sense of the multifarious “materialities of information” (Dourish, 2017) and related practices, it is necessary to weave together a variety of theoretical points of view into a “methodological criss-crossing” (Souchier *et al.*, 2003: 29-30). Each perspective may thereby complement the others in a reflexive and critical manner. The research plan is structured around the criss-crossing of three disciplinary perspectives, with the concept of space acting as the cornerstone: (RO 1) a comparative history of the birth and development of both platforms will aim at recovering the political dimension of digital libraries; (RO 2) a media study of their architectures, interfaces, and codes will help understand them as mediatic spaces; (RO 3) an ethnographical inquiry, both qualitative and quantitative, of their users’ practices will flesh out the platforms as experiential spaces.



In a continuous effort to engage reflexively both with my research object and my own use of digital tools “behind-the-scenes” (Hargittai and Sandvig, 2015), I will sustain an epistemological critique of my mixed methodology. Each perspective shedding light on the assumptions and blind spots of the others, they dovetail into an organic method endowed with its own consistence.

First, “traces” and their hermeneutic are traditionally construed as paradigmatic concepts for the epistemology of humanities and social sciences (Ginzburg, 1979).<sup>7</sup> But while the hermeneutic of “analogue” traces would reveal worlds beyond their seeming insignificance, digital traces would appear to make sense only when heavily processed in vast amounts. The

<sup>7</sup> With the advent of digital traces, debates have resurfaced on whether individuals leave traces on their path or if they are constituted *a posteriori* by them (notably Galinon-Mélenec, 2011; Galinon-Mélenec and Zlitni, 2013; Galinon-Mélenec *et al.*, 2015).



historiographical (RO 1) and semiotic (RO 2) critique of digital traces and of their quantitative processing as a hermeneutic (RB 3.3) will act as a critical safeguard in their manipulation.

Second, an understanding of digital libraries as spaces, and navigation thereof, should not pass uncriticized. Acknowledging the reification—or *effet de réel*—produced by visual representation (Drucker, 2014, 2020),<sup>8</sup> digital libraries cannot be taken as uncharted territories only waiting to be mapped (Rodighiero, 2021). Shuttling back-and-forth between qualitative (RBs 3.1 and 3.2) and quantitative (RB 3.3) ethnography, the digital could be conceptualized as a milieu (Bachimont, 2015; Le Deuff, 2016), and the Web in particular as an “information milieu” (Ghitalla *et al.*, 2003), generated by navigational practices themselves, as well as their visual representation. While the Internet dynamic has early been qualified as an “ecology of knowledge” (Huberman, 2001), only recently has this perspective been developed (e.g., Szoniecky, 2018). This caveat could contribute to a renewed ecological approach to knowledge in the making (Dumas Primbault *et al.*, 2021a).

Finally, the methodological criss-crossing at the heart of this project will shed new light on the so-called contemporary “episteme”—either “computation” in Golumbia (2009), “control” in Franklin (2015), or “software” in Sack (2019). By confronting the material conditions of possibilities of navigation (RO 2) with its actual material practices of appropriation (RO 3), within specific social and cultural contexts (RO 1), this project strikes a middle ground between technological determinism and absolute constructivism—which can respectively be construed as artefacts produced by single methodologies either too focused on technical objects or on individuals’ discourses (Souchier *et al.*, 2003: 30). Building on a conception of the digital as a technical milieu and drawing on the recent reappraisal of Simondonian philosophy of technology (e.g., Rieder, 2020), navigation appears as the dialectical coupling of user and technology through layers of appropriated mediations.

### *Research Objective 1: Cross-Histories of two Digital Libraries*

Celebrating its 30<sup>th</sup> birthday this year, arXiv was established in 1991 at the Los Alamos National Laboratory (LANL) by physicist Paul Ginsparg. It initially started as an email service and an FTP repository before it became a website in 1993. Designed as an open-access platform for electronic pre-prints in TeX format for the field of physics, arXiv.org was also strongly involved in the early Open Archives Initiative Protocol Development. Hosted and funded by Cornell University since 2001, it provides open access to 2 million scientific e-prints in fields ranging from physics and mathematics to computer science. Spin-offs have sprouted since: beta platforms to be personalised (iarXiv), or copycats for other disciplines (bioRxiv, medRxiv, socaRxiv).

Founded in 1997, Gallica is the online platform of the Bibliothèque nationale de France—the second largest library in the world—dedicated to providing free access to digitized historical sources preserved at BnF. Thought from the beginning as a “virtual library” it was endowed with a patrimonial mission. Continuously developed over the following decades, it now gathers almost 7 million documents, in 8 languages of diverse types ranging from books, manuscripts, newspapers, to photographs, and musical scores. It is widely used within the francophone community of humanities and social sciences. In October 2021, the BnF inaugurated a DataLab dedicated to the digital humanities.

Work Package 1: address the long tail of the history of information management by proposing a cross-history of the birth and development of arXiv and Gallica. By situating them in the institutional landscape, the technological background, and the cultural policies that led to their creation, either for patrimonial purposes, or to foster open science, this historical perspective will flesh out digital libraries as political spaces.

The first source material available is the administrative documentation archived by their respective institutions: the Bibliothèque nationale de France for Gallica;<sup>9</sup> the Los Alamos National Laboratory<sup>10</sup> and Cornell University<sup>11</sup> for arXiv. The administrative perspective will help me identify the actors—both individual and institutional—involved, their respective

<sup>8</sup> Particularly with topological data analysis (see RB 3.3) as it is a quantitative method that may produce qualitative results in the study of shape.

<sup>9</sup> The most recent “Archives institutionnelles de la Bibliothèque” are preserved on the François-Mitterrand site (<https://archivesetmanuscrits.bnf.fr/pageCollections.html>), the archives of the Ministère de la culture are held at the Archives nationales.

<sup>10</sup> The reports of the LANL are held at its Research Library (<https://www.lanl.gov/library//index.php>).

<sup>11</sup> Official university records, office records, and research records are held at the Cornell University Archives (<https://rare.library.cornell.edu/preserving-cornell-university-history>).

objectives and the negotiations that might have ensued, the choice of technical media through the discussion of nascent standards and protocols, as well as early prototypes and subsequent versions. The second source material available are the actors themselves. An oral history based on interviews with founders, developers, librarians, early users, and volunteer community moderators (in the case of arXiv) should allow me to put into critical perspective the administrative archive. Thirdly, as Internet objects, platforms bring forth the question of the Web archive, both through the issues raised during their development and through traces of their historical versions they may have left behind (Schafer, 2018).<sup>12</sup> This Web-archival history will also look at other types of sources, such as online videos,<sup>13</sup> and reflect on the possibility to deal with email correspondences.

The research blocks are organized by topics.

#### Research Block 1.1: Governance

While both platforms were born amid the cyberspace dream, the reappraisal of the notion of space is important as it reintroduces the political dimension of these platforms, inseparable from their technological dimension. The co-presence of a wide variety of entities and actors (institutions, users, authors, publishers, beneficiaries, stakeholders) raises the issue of the governance of this political space: who are these entities, how do they interact, with what objectives, and through which negotiations is it ensured that the demands of all actors are met (Mounier, 2020). Paying close attention to their differences in scopes, designs, and developments, or, on the contrary, to what they have in common—and would appear to be peculiar to the digital—, this research block is dedicated to unravelling the political spaces of arXiv and Gallica as an intertwining of different modes of governance— institutional, academic, community—or reactions to external modes of governance perceived as disruptive,<sup>14</sup> resulting in the structuration, hierarchization, or invisibilization of content, in the choice of language(s) and technical standards.

#### RB 1.2: Communities

Contrary to other infrastructures where stakeholders and clients are strictly distinguished, the specificity of knowledge infrastructures such as digital libraries is that they host and provide access to commons that may be of use to all involved parties. Therefore, specific attention will be given to the many non-mutually exclusive communities—researchers as users, moderators, contributors, board members...—involved in the development, maintenance, and use of arXiv and Gallica, and how they structure the governance of commons by ensuring the sustainability of their access and circulation.

#### RB 1.3: Open access

Beyond specific communities, both the platforms studied were designed from the beginning as open access repositories. This research block will inquire about what flavours of open access were developed by arXiv and Gallica: what legal solutions were found to issues of intellectual property, what archival standards they adopted or developed, through which protocols, how they guarantee interoperability. Thereby taking a strong stance in the landscape of scientific publishing and distribution, an attention to press coverage could be stressed, shedding light on reception both in scholarly journals and among the wider public.

### *Research Objective 2: Digital Libraries as Technical Media*

Both platforms present themselves as websites accessible to the users through the window of their Internet browsers. arXiv displays a rather simple, linear, and somewhat “lo-fi” homepage diving right in disciplines and subdisciplines, while offering the possibility to quickly see either “new” or “recent” contributions. On the contrary, Gallica opens on an assemblage of blog posts and topical archival selections; below are tutorials and school material. The multiple entry points into its collections—text search, document types, topics, and geographical areas—are situated in a thin bar on top of the window. To better understand how political spaces find material shape in the form of digital objects, we must consider them as fully-fledged technical media.

---

<sup>12</sup> Early versions of Gallica are available on the Wayback Machine and can be used to trace changes in its self-fashioning discourses, or in its interface and architecture (see RO 2). Although arXiv has been recorded in the past, it decided to ban archival robots and erase records. While some traces are scattered in papers and websites, this refusal to be archived deserves some interest.

<sup>13</sup> The repository’s 30<sup>th</sup> birthday was the occasion for its current actors to reflect in Youtube videos on its past and think about its future (see, for example, <https://www.youtube.com/watch?v=12peSzGwLBO>).

<sup>14</sup> As an example, Jean-Noël Jeanneney, president of the BnF from 2002 to 2007, published a book in 2005 fashioning Gallica as the French public response to the creation of Google Books.

Work Package 2: provide a functional description of arXiv and Gallica construed as technical objects mediating between humans and computers, by unravelling the many layers of mediation that allow for communication processes and interaction with digital content. By focusing on how these layers of mediation may frame specific regimes of attention on the part of users, this will shed light on how researchers may, or may not, navigate digital libraries. Indeed, that neither Gallica nor arXiv have built-in recommendation systems is at once a curse and a blessing: by default, they are not dedicated to navigation and exploration—they would at first seem rather made for targeted search and information extraction—but it also means that if one succeeds in modelling relevant navigation paths, then they are not documenting the recommendation algorithms but the very practices of the users. However, this does not mean that there is no “interface effect” (Galloway, 2012; Hookway, 2014) in arXiv and Gallica.

The research blocks are organized in a descending fashion from surface to deeper layers of mediation.

#### Research Block 2.1: Interface and structure

The first mediation the user is exposed to is the surface of the graphical user interface (GUI). The GUI allows for general navigation within the structure of the website that organizes the many functions of the platforms—homepage, news, search, results, documents, contact, help, and blog. The semiotic study of both interface and structure (e.g., Pignier, 2009; Reyes, 2019), paying attention to interactive signs (or “*signes passeurs*” such as links, fields, lists, hover, scrollbar...) and their effects, to how simple and advanced searches can be carried and how search results can be formatted (lists, tables, graphs...) will highlight the possible affordances users may seize, follow, or hack when manipulating these medias. This RB could be complemented with a quantitative analysis of clicks within pages (e.g., Lamprecht, 2015) if coupled with RB 3.3. Besides the user interface, both platforms also feature APIs structured on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). These backend interfaces allow for the communication between machines and will be used to enrich server logs in RB 3.3.

#### RB 2.2: Information architecture

Beneath the website’s functional structure lies its information architecture allowing for the sorting, retrieval, and hierarchization of content through simple or advanced searches. Metadata structures and classification standards enact specific orders of knowledge that will frame the users’ paths (Bowker and Leigh Star, 1999; Sarrazin *et al.*, 2019). Because both libraries follow the OAI protocol, they are constrained to make use of the vocabulary provided by the Dublin Core Metadata Initiative (DCMI) consisting in fourteen descriptors comprising notably creator, date, title, publisher, and format. One important metadata element for this research project is the subject or “topic of the resource”. arXiv uses an *ad hoc* partition of research fields in disciplines, subdisciplines, and possible specialisations resulting from the history of the platform itself. Subject classes are not mutually exclusive: authors have to specify the core discipline but can add other subsidiary classes. Gallica on the other hand relies on the Dewey Decimal Classification (DDC) it inherited from the BnF as all the online documents are digitisations of physical ones classified according to the library’s system. Therefore, disciplinary descriptors are mutually exclusive and limited to the tree-like structure of the Dewey classes. Beyond documents metadata, matters of file formats, markup languages, and possible web ontologies to structure databases will help understand how documents can be manipulated and the platforms navigated.

#### RB 2.3: Code

Finally, following Cardon’s call to “lay algorithms on the couch” so as to perform a “critical radiography” (Cardon, 2015: 12-13), a deeper layer of mediation can be analysed by opening the black box of search engines and understand the mechanisms at work behind the hierarchization of proposed content. Whether their code is in open access or by attempting to retro-engineer their main features, this will shed light on how variables, search results, filtering, or highlighting may guide ways to navigate through the platforms, whether they take the users to other fields, or even to other platforms. Furthermore, while Gallica has recently been testing an advising system for curious documents, independent arXivLabs projects such as “CORE Recommender” or “Connected Papers” are currently developing opt-in recommending systems. Scrutinizing their algorithmic specifics is necessary to understand what kind of navigation they embody. Reciprocally, the present research could benefit these endeavours (see Relevance and Impact section).

### *Research Objective 3: An Ethnography of Digital Navigation*

As mere affordances of arXiv and Gallica, their interfaces, architectures, and codes only provide the conditions of possibility for the actual navigation of digital libraries. In order to bridge the gap between the technical media and their practical appropriation, or even hacking, I will need to shift focus from the platforms themselves to their users, and more specifically to how they materially interact with affordances within the experiential space of practices. Building on the anthropological postulate that we can conceive of groups of researchers as epistemic cultures (Knorr Cetina, 1999), *i.e.*, communities sharing values, norms, tools, and spaces, ethnographic methods can be employed to document their practices. Striving not to over-emphasize differences between or coherences within distinct epistemic cultures in order to avoid former stereotypes (e.g., that only for humanists would the journey be part of the research), a broader view of research practices can be depicted through interviews and non-participant observation of users. This can be achieved by using mixed methods from (digital) ethnography (Venturini *et al.*, 2014), combining quantitative “big data” with qualitative “rich data” (Snee *et al.*, 2016).

Work Package 3: provide a thick description (Geertz, 1973) of navigational research practices as both observed by the ethnographer and objectified by the practitioners themselves. By using different tools (semi-directed interviews, non-participant observation, and quantitative topological analysis) for different levels of analysis, this thick description should bring forth sets of multifarious practices and media, shed light on how they coexist, sometimes in an undifferentiated manner so much so that it becomes difficult to disentangle something as “the digital”, and help identify phases of the research and regimes of navigation.

The research blocks are organized by methodology.

#### Research Block 3.1: Semi-directed interviews

First, about 25 users per platform will be recruited, of all age above 18, who do research either professionally or as their principal activity, with connection times to the platform superior to 10mn. They will each be submitted to one semi-directed interview of 1 to 2 hours aimed at the objectification of their research practices, and navigation in particular. Interviews will be recorded and transcribed. Particular care will be taken in conducting the interviews online, reflecting on the methodological conditions of such a set-up (e.g., Theviot, 2021).

After establishing a basic socio-professional profile (age, country, professional status, discipline(s), research topic and method, primary sources, secondary literature, and data used), the interview will proceed in four concentric steps gradually going from the platform of inquiry towards broader information practices and navigation: i) within the platform of inquiry, ii) cross-platform practices, iii) digital practices, iv) all information research practices in general. In steps i) and ii), by progressing from the initial connection to the platform towards longer sessions, interviewees will be led to express how they link their documents practically (through new search, opening tabs, refining search, using suggestion...) and intellectually (association of idea, references in bibliography, people with whom they talked). Steps iii) and iv) will address other practices of live reading, notetaking, referencing, downloading, sorting, and how navigation is integrated within a supposed “processing chain”. Specific attention will be paid to the degree of easiness demonstrated by the interviewees in the objectification of their own practices, and to the metaphors or other figures of speech they use to describe them.

The interview will be followed by around 30 minutes when interviewees will be asked to share their screen and perform three tasks (Ghitalla *et al.*, 2003: 198): find a specific document with partial metadata (information-mining), constitute a small bibliography on a given subject (information-farming), and wander freely following their interests and curiosity of the moment (information-strolling). The tasks are aimed at revealing tacit details or unconscious habits. Particular attention will be paid to how interviewees make use of the interface’s affordances. Moreover, the discrepancy between discourse and performance, rather than inconsistency, is important evidence to assess the practical reflexivity of users.

#### RB 3.2: Non-participant observations

Second, out of the 25 respondents per platform, 5 will be selected on a voluntary basis as a representative sample and will be subjected to repeated semi-directed interviews of no more than 1 hour per month (to be defined with each respondent). Interviewees will be asked to record one hour of their research, whenever they want, to be shared with the interviewer. The

video, and its comparative analysis with other research blocks in progress, will serve as the basis for the monthly discussion. This hybrid setup, mixing observation with an interview centered around it, is aimed at identifying potential diachronic phases in research, and allowing to trace cross-platform and cross-software practices.

I will also carry a more thorough on-site non-participant observation of these selected 10 respondents in total. For at least half a day, depending on what we agree on, I will observe them at one of their places of work of their choice so as to document through notes and documents the on-site materiality of their information practices: if they have access to a library, how they store and organize their documents, whether they have experimental apparatuses, when and where they can discuss with colleagues. Non-participant observation will also allow me to document how different media are woven together in practice, in particular what researchers do with the documents they found while navigating—their “processing chain”—: whether they read them in line, download, print, or modify. On-site discussion should also allow me to assess what changed since the lock-in.

### RB 3.3: A topological analysis of server logs

Parallel to interviews and observations, I will conduct the quantitative analysis of server logs with TDA.<sup>15</sup> By analogy with the paths followed by researchers within the physical space of a library, I will endeavour to model and extract individual navigational paths from the server logs of arXiv and Gallica (steps 1 and 2). Using TDA, I will then be able to *construct* the “space” of disciplines (or subject classes), that is, rather than a fixed territory inferred from link structures or bibliometric indices, a *representation* of the conceptual proximity of fields of study as generated by the navigational practices of users themselves (step 3). Eventually, drawing the navigation paths within that representation space will allow me to study their morphology and cluster them according to their topological features—either reduced to single points, or circular, star-shaped, or tree-like paths (steps 4 and 5).

This research block will have to be undertaken in close connection with RBs 3.1 and 3.2 as only the confrontation of the quantitative results with the qualitative part of the study will tell us what each cluster might mean—whether it can be understood as a strategy, linked with phases of research, or simply discarded. These findings will serve as probes to lead discussions in interviews. Reciprocally, the interviews and observation of the respondents will help refine the models.

#### *Step 1 – Gathering Data: Harvest Server Logs from the Two Platforms*

As a preliminary step, I will contact Gallica and arXiv to request individualized anonymous logs. It is essential I anticipate the need for the approval of EPFL’s ethical committee before I renew the existing convention with Gallica<sup>16</sup> and sign one with arXiv.<sup>17</sup>

#### *Step 2 – From data to corpus: model and extract navigational paths from server logs*

The next step will be to model and extract from server logs a coherent corpus of “navigational paths”. As a first approximation, a linear model of series of three or more non-redundant documents consulted within 60mn of each other already proved relevant for Gallica. More complex non-linear models can be informed by users’ interviews, considering arborescent ramification when the next request is made from an earlier page; distinguishing post-query visits by modelling the search result pages as pivots; accounting for flicking or zooming... Reciprocally, models will be tested in interviews. Furthermore, this will allow me to estimate the basic features of navigation practices (frequency, length, time).

Note that extracted paths will be enriched with relevant metadata garnered through the platforms’ APIs. Moreover, on platforms whose corpora gather a few million items, the scale of analysis cannot be the single document. It will therefore be necessary to reduce granularity. For Gallica, the sole Dewey class proved to be relevant. It is also possible to refine with the publication year, the type of document, or other metadata.

---

<sup>15</sup> Reciprocally, the Web conceived of as a “laboratory of culture” offers new mathematical challenges for applied topology, and in particular for TDA. In the continuation of the CROSS project, this could be the basis for a collaboration focused on the mathematical and computational technicalities of this research block with EPFL’s Laboratory for Topology.

<sup>16</sup> The current agreement between EPFL and Gallica covers April 2016 – April 2017 and will be renewed with more recent and bigger data.

<sup>17</sup> The platform already gathers “anonymized browsing logs” and “pathways navigated through the website” that it shares with trusted third-party services. They provide institutions with an “arXiv bulk data” service to work on their data.

## METHODOLOGY FOR AN ETHNOGRAPHY OF THE DIGITAL

### • Semi-directed interviews

- 1–2 hours
- c. 25 interviewees per platform

  1. Basic sociological profile
  2. Within one platform
  3. Between platforms
  4. On the computer
  5. Outside the computer
  6. Observing three tasks

### • Repeated semi-directed observations over 2 years

- ≤ 1 hour per month
- c. 5 respondents per platform

  - Discussion of 1 hour of recorded research
  - Phases of research
  - Cross-platform and cross-software practices

### • Prolonged non-participant observation at the workplace

- ≥ 1/2 day
- c. 5 respondents per platform

  - On-site informational practices
  - Intertwining of multiple practices and media
  - Processing chain

Observation of practices to models

Probes to lead discussion

## PIPELINE FOR A DIGITAL ETHNOGRAPHY

### 1. Harvest server logs

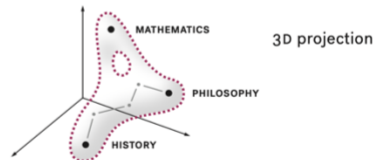
```
##gacd1db099decb93848d##France##Angers
      IP ADDRESS      COUNTRY  CITY
##--[24/Nov/1990:10:30:25] "GET /index.html"
      TIMESTAMP      REQUEST
HTTP/1.1 500 149 "http://savoirs.app"
      PROTOCOL CODE SIZE REFERENT
```

### 2. Model and extract navigational paths

Path = (doc<sub>1</sub>, doc<sub>2</sub>, doc<sub>3</sub>, ...)

metadata<sub>1</sub>

### 3. Generate a topological space of disciplines



### 4. Cluster paths by morphological features



### (5. Refine clusters by sizes and temporalities)

### Step 3 – Generate a topological space of disciplines

To study the morphology of navigational paths, it is necessary to embed them within a mathematical metric space that will quantify the conceptual proximity between disciplines and allow for their visual representation accordingly. By representing navigational paths as graphs (weighted by frequency and directed by timestamps), TDA algorithms will allow me to generate a topological space from the aggregation of such graphs—a simplicial complex. The meaning of this object will have to be thoroughly discussed. While traditional “maps of science” work from citations and hyperlinks (Börner *et al.*, 2012) or from the textual features of abstracts (Dias *et al.*, 2017)—thereby drawing a cartography without its users—, this topological space will be generated out of server logs, representing how researchers navigate platforms and orient themselves with their metadata. Drawn from the point of view of users, and non-existent without them, it is more akin to a representation of a *practical* order of knowledge, rather than a map, and acts as a counterpoint to taxonomies in the forms of trees, graphs, and diagrams.

Acknowledging that no definitive “map” of a supposed territory can be drawn and that only a comparative approach to many perspectives can be informative (Gläser *et al.*, 2017), it will be necessary to emphasize that the visual artefacts produced by the projections of these high-dimensional spaces are platform-specific, therefore highly dependent on subject classes (RB 2.2), and generated by practices, therefore continuously redrawn. Thus, their consultation should be made interactive, and their readers be left the freedom to change the projection method or adjust the chronological boundaries, allowing for the reconstitution of their dynamic through series of snapshots.

### Step 4 – Model paths as walks within the topological space and cluster them by morphology

Finally, individual paths will then be embedded as vectors within a high-dimensional Euclidean space whose metric, defined by a selected number of topological features, will account for their morphological resemblance, *i.e.*, the pairwise distance between paths will measure their morphological dissimilarity. The clustering of vectorized paths (with a dedicated method still to be defined according to the embedding) will help identify several patterns in navigational practices: points (*i.e.*, single disciplines), pairs, and circular, starlike, or more complex shapes.<sup>18</sup> The shapes of said patterns will then have to be interpreted in light of the

<sup>18</sup> Additionally, clustering could be refined, for example, into subclusters computed by comparing the lengths of disciplinary leaps or taking into account temporal dynamics (the length of vertices encoding timestamps)—this could be achieved with Dynamic Time Warp (DTW).

qualitative analysis: e.g., the trivial cluster (one dot within one single discipline) can represent either someone who did a very directed search (short path), or someone who's constituting a corpus within one very specific discipline (longer path without leaps). Not a taxonomy, clusters will rather denote various regimes of navigation that are neither found pure nor mutually exclusive and may depend or not on deliberate strategies.

## Schedule and milestones

- **Research articles (RA)** both in specific fields and in interdisciplinary journals/platforms all with open access ensured by EPFL. Other forms of digital publication will be privileged, such as pre-prints.
- Presentation in **international conferences (IC)**, with printed proceedings, on digital libraries (such as ICDLM, ICCEDL or TPDJ), human-computer interaction (CHI), digital humanities (ADHO, EADH), and STS (4S, EASST).
- At least two interdisciplinary **workshops (WS)** will be organized.
- All the **code (C)** produced during the project will be made public in open access and open source (GNU license) on the laboratory's Github with relevant documentation.
- In accordance with the FAIR principles, **data (D)** will be made public and open as much as possible.
  - In their consent form, interviewees will specify whether they agree to their video/only the audio/only the anonymized transcript/nothing be made publicly available.
  - The agreements signed with the platforms will set matters of open server logs.
- A novel **course** in “Computational Social Sciences” in the DH master of EPFL, experimenting with collaborative online platforms to initiate discussion and debate with the students.

### WP1: Cross-Histories of Digital Libraries (Year 1)

Anticipate ethical clearance (**EC**) for interviews of RB 1.2 (indicated in red on the Gantt chart)

**RA1** for, e.g., *portal* or the *International Journal on Digital Libraries*

### WP2: Digital Libraries as Technical Media (Year 2)

**WS1** on Human-Computer Interaction

**RA2** for HCI (e.g., *TOCHI* or *Human-Computer Interaction*) and **RA3** for media studies (e.g., *Media, Culture, & Society*)

### WP3: Ethnography of Digital Navigation (Years 3 and 4)

Anticipate ethical clearance (**EC**) for interviews (RB3.1), observation (RB 3.2), data acquisition and processing (RB 3.3) and agreements (**A**) with platforms for data acquisition (indicated in red on the Gantt chart)

**WS2** on Computational Social Sciences

**RA4** and **RA5** for, e.g., *Réseaux*, *Learned Publishing*, or the *Journal of Association for Information Science and Technology*

### Reflexive critique: on the “methodological criss-crossing”

**RA6** for, e.g., *DHQ*, or *Theory, Culture, & Society*, or *Social Science Computer Review*

		Year 1				Year 2				Year 3				Year 4									
Quarters		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4						
WP1 Cross-Histories of Digital Libraries	RB 1.1	█																					
	RB 1.2	█ EC	█																				
	RB 1.3	█		█		█		█															
WP2 Digital Libraries as Technical Media	RB 2.1					█		█															
	RB 2.2					█	█		█														
	RB 2.3					█		█		█													
WP3 Ethnography of Digital Navigation	RB 3.1									█ EC & A													
	RB 3.2									█				█									
	RB 3.3									█				█									
<i>Basso continuo</i> Epistemology of Digital Research		█																					
Deliverables						RA1				RA2				RA3				RA4				RA5 & RA6	
										WS1										WS2		C&D	

## Relevance and impact

This research partakes in the broader endeavour to study and design an ever more needed “scholarly information infrastructure” (Borgman, 2007). More specifically, the focus on navigation and exploration may help revive the early “political force of the Internet” (Auray, 2011: 330) toward a renewed “democratization” (Cardon, 2010). Practically, four leads in that direction could be explored: i) extend the study to public platforms; ii) give users their data back; iii) involve users in the design of infrastructures; iv) design a digital compass.

First, at the heart of today’s necessary reform of the digital, the issue of citizen participation in science is key to a greater and freer digital navigation that would escape the centralisation and standardisation of information content and practices. Shifting the focus from researchers’ practices within academic platforms towards more public interfaces such as Wikipedia (Barbe *et al.*, 2015, 2021; Tkacz, 2015), we could address this issue raised by the navigation of lay audiences on open platforms. On the side of information and communication studies, this further comparative study would inform us on public practices of knowledge consultation and production—on citizen science in other words—, or simply on common practices ranging from intellectual musings to curious strolls.

Second, stemming from the conviction that the data generated by users while navigating belong to them, the tools developed during this research could be directly integrated within the studied platforms for users to assess and control their personal “digital presence” (Merzeau, 2010a) in the form of an interactive history. On an opt-in basis only, users could have easy access within their personal account to their own logs and, through a dedicated interface (RB 2.2), to a set of analytics and data visualization tools to process the logs according to their own parameterization. They could therefore have direct access to the interactive landscape of knowledge inferred from aggregated paths (RB 3.3), as well as generate one from their own interaction traces only. They could visualize their navigation along aggregated mean paths or beaten tracks, thereby learning how to “inhabit” the digital ecology (Merzeau, 2010b).

Third, in an effort to involve users in the development of the infrastructures they need, this research specifically led from their perspective should be of great use. Supplementing the descriptors of taxonomies, tags could be inferred when two disciplines are usually navigated together. Such a folksonomy would surely facilitate browsing (Auray, 2007) as tags are mutually non-exclusive and offer the possibility to evolve through time in a more flexible manner than taxonomies (RB 2.1). Reciprocally, researchers’ online behaviour could be used to design the layout of physical libraries to foster exploration, therefore embodying the dream of a collective Warburg library.

Finally, going beyond classical bibliometrics, in order not only to measure, but more importantly to facilitate and foster interdisciplinarity, our research could feed the scholars’ practices back by providing them with a tool to orient themselves in these multidimensional topological spaces—a digital compass. Allowing them to fine-tune the parameters of a path-generating algorithm, they would then be able either to follow pre-framed typical pathways (e.g. for novices willing to see what a theoretical physicist would consult<sup>19</sup>), or to calculate average paths according to a set of variables (e.g. a weighted combination of disciplines for a given level of study); either to plot a course along a “scenic route” made for intellectual musing, or to venture off the beaten track along unprecedented trajectories, through byways and crosscuts. Either with traditional mathematical metrics (Leibon and Rockmore, 2013) or taking advantage of recent TDA development undertaken at EPFL on trajectory prediction on topological spaces with neural networks (Ebli *et al.*, 2020), this compass could lead users towards the less consulted areas of digital libraries, or to leap from platform to platform making them interoperable.

This could be achieved by applying for example to a SNF Eccellenza Grant, or an ERC Starting Grant. Collaborations on applied R&D projects could be led with Innosuisse/SNF grant BRIDGE Proof of Concept or through international partnerships thanks to Innosuisse grant EUREKA.

---

<sup>19</sup> The educational virtues of “user-transaction-based recommendation strategy” in digital libraries have been demonstrated by Kortemeyer and Droschler (2021).



## Bibliography

- Abbott, Andrew. (2014). *Digital Paper: A Manual for Research and Writing with Library and Internet Materials*. Chicago: The University of Chicago Press.
- Anderson, Theresa D. (2005). "Relevance as Process: Judgments in the Context of Scholarly Research". *Information Research* 10(2). <http://www.informationr.net/ir/10-2/paper226.html>
- Anderson, Theresa D. (2006). "Uncertainty in Action: Observing Information Seeking within the Creative Processes of Scholarly Research". *Information Research* 12(1). <http://informationr.net/ir/12-1/paper283.html>
- Antonijević, Smiljana. (2015). *Amongst Digital Humanists: An Ethnographic Study of Digital Knowledge Production*. New York: Palgrave MacMillan.
- Antonijević, Smiljana. (2020). "Digital Workflow in the Humanities and Social Sciences: A Data Ethnography". In: Jerome W. Crowder, Mike Fortun, Rachel Besara, and Lindsay Poirier (eds.). *Anthropological Data in the Digital Age: New Possibilities – New Challenges*. Cham: Palgrave MacMillan, 59-83. [https://doi.org/10.1007/978-3-030-24925-0\\_4](https://doi.org/10.1007/978-3-030-24925-0_4)
- Antonijević, Smiljana, and Ellysa Stern Cahoy. (2014). "Personal Library Curation: An Ethnographic Study of Scholars' Information Practices". *portal: Libraries and the Academy* 14(2), 287-306. <https://doi.org/10.1353/pla.2014.0010>
- Antonijević, Smiljana, and Ellysa Stern Cahoy. (2018). "Researcher as Bricoleur: Contextualizing humanists' digital workflows". *Digital Humanities Quarterly* 12(3). <http://www.digitalhumanities.org/dhq/vol/12/3/000399/000399.html>
- Auray, Nicolas. (2007). "Folksonomy: A New Way to Serendipity". *Communications and Strategies* 65, 67-91.
- Auray, Nicolas. (2011) "Les technologies de l'information et le régime exploratoire". In: Van Anel, Pek, and Danièle Boursier (eds.). *La sérendipité. Le hasard heureux*. Paris: Hermann, 329-343. <http://ses-perso.telecom-paristech.fr/auray/2010AuraySerendipite.pdf>
- Bachimont, Bruno. (2015). "Le numérique comme milieu : enjeux épistémologiques et phénoménologiques. *Principes pour une science des données?*". *Interfaces numériques* 4(3), 385-402. <https://doi.org/10.25965/interfaces-numeriques.386>
- Barbe, Lionel, Louise Merzeau, and Valérie Schafer (eds.). (2015). *Wikipédia, objet scientifique non identifié*. Nanterre: Presses Universitaires de Paris Nanterre. <https://doi.org/10.4000/books.pupo.4079>
- Barbe, Lionel, and Marta Severo (eds.). (2021). *Wikipédia, objet de médiation et de transmission des savoirs*. Nanterre: Presses Universitaires de Paris Nanterre. <https://doi.org/10.4000/books.pupo.15040>
- Bates, Marcia J. (1989). "The design of browsing and berrypicking techniques for the online search interface". *Online Review* 13(5), 407-424. <https://doi.org/10.1108/eb024320>
- Bates, Marcia J. (2007). "What is browsing—really? A model drawing from behavioral science research". *Information Research* 12. <http://www.informationr.net/ir/12-4/paper330.html>
- Beaudouin, Valérie, and Jérôme Denis. (2014). "Observer et évaluer les usages de Gallica. Réflexion épistémologique et stratégique". [Research Report] BnF; Télécom ParisTech. <https://hal.archives-ouvertes.fr/halshs-01078530>
- Beaudouin, Valérie, Isabelle Garron, and Nicolas Rollet. (2016). "Je pars d'un sujet, je rebondis sur un autre' : Pratiques et usages des publics de Gallica". [Research Report] Télécom ParisTech; BnF; Labex Obvil. <https://hal.archives-ouvertes.fr/hal-01709238>
- Bert, Jean-François, and Jérôme Lamy. (2021). *Voir les savoirs. Lieux, gestes et objets de la science*. Paris: Anamosa.
- Bilenko, Mikhail, and Ryen W. White. (2008). "Mining the Search Trails of Surfing Crowds: Identifying Relevant Websites From User Activity". *WWW 2008*, 51-60. <https://doi.org/10.1145/1367497.1367505>
- Blair, Ann, Paul Duguid, Anja-Silvia Goeing, and Anthony Grafton (eds.). (2021). *Information: A Historical Companion*. Princeton: Princeton University Press.
- Borgman, Christine L. (2000). *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. Cambridge, MA: MIT Press.
- Borgman, Christine L. (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press.
- Borgman, Christine L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.
- Börner, Katy, Richard Klavans, Michael Patek, Angela M. Zoss, Joseph R. Biberstine, Robert P. Light, Vincent Larivière, and Kevin W. Boyack. (2012). "Design and Update of a Classification System: The UCSD Map of Science". *Plos One* 7(7). <https://doi.org/10.1371/journal.pone.0039464>
- Boukacem-Zeghmouri, Chérifa. (2012). "Online Article Searching on Publisher Platforms by STM French Scholars: Findings and Analysis." *Canadian Journal of Information and Library Science* 36(3-4), 88-105.
- Boullier, Dominique, and Jacques Lévy. (2016). "Topographies/topologies. Langages spatiaux, spatialités, espaces". Special Issue of *Réseaux* 195(1). <https://doi.org/10.3917/res.195.0009>
- Bowker, Geoffrey C., and Susan Leigh Star. (1999). *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.
- Cardon, Dominique. (2010). *La démocratie Internet. Promesses et limites*. Paris: Seuil.
- Cardon, Dominique. (2015). *À quoi rêvent les algorithmes. Nos vies à l'heure des big data*. Paris: Seuil.

- Case, Donald O. (2002). *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. San Diego: Academic Press.
- Chartron, Ghislaine, Benoît Epron, and Annaïg Mahé (eds.). (2010). *Pratiques documentaires numériques à l'université*. Villeurbanne: Presses de l'ENSSIB. <https://doi.org/10.4000/books.pressesensib.1162>
- Citton, Yves. (2017). *Médiarchie*. Paris: Seuil.
- Daston, Lorraine. (ed.). (2017). *Science in the Archive: Pasts, Presents, Futures*. Chicago: University of Chicago Press.
- Da Sylva, Lyne and Pascal Cuxac (eds.). (2017). "Analyse et exploitation des bibliothèques numériques". Special issue of *Document numérique* 20(2-3). <https://doi.org/10.3166/dn.2017.00015>
- Deveaud, Romain. (2018). "Modalités d'accès au savoir ouvert sur les plateformes d'OpenEdition". Aix Marseille Université.
- Dias, Laércio, Martin Gerlach, Joachim Scharloth, and Eduardo G. Altmann. (2017). "Using text analysis to quantify the similarity and evolution of scientific disciplines". *Royal Society open science* 5. <https://doi.org/10.1098/rsos.171545>
- Dourish, Paul. (2017). *The Stuff of Bits: An Essay on the Materialities of Information*. Cambridge, MA: MIT Press.
- Drucker, Johanna. (2014). *Graphesis: Visual Forms of Knowledge Production*. Cambridge, MA: Harvard University Press.
- Drucker, Johanna. (2020). *Visualization and Interpretation: Humanistic Approaches to Display*. Cambridge, MA: MIT Press.
- Dumas Primbault, Simon. (2020a). "Le compas dans l'œil. La "mécanique géométrique" de Viviani au chevet de la coupole de Brunelleschi". *Revue d'histoire des sciences* 73(1), 5-52. <https://doi.org/10.3917/rhs.731.0005>
- Dumas Primbault, Simon. (2020b). "An Ink-and-Paper Automaton: The Conceptual Mechanization of Cognition and the Practical Automation of Reasoning in Leibniz's *De Affectibus* (1679)". *Society and Politics* 13(2), 87-113. <http://socpol.uvvg.ro/docs/2019-2/6.%20Dumas.pdf>
- Dumas Primbault, Simon. (2020c). "Un milieu d'encre et de papier. Brouillons, notes et papiers de travail dans l'archive personnelle de Vincenzo Viviani (1622-1703)". *Cahiers François Viète* III(10). [https://cfv.univ-nantes.fr/medias/fichier/dumasprimbault-cahierscfv-serie-iii-10\\_1618390611224-pdf?ID\\_FICHE=1415775](https://cfv.univ-nantes.fr/medias/fichier/dumasprimbault-cahierscfv-serie-iii-10_1618390611224-pdf?ID_FICHE=1415775)
- Dumas Primbault, Simon. (2021). "Leibniz en Créateur. Un regard matériel sur sa machine à calculer". *Savoirs*, forthcoming.
- Dumas Primbault, Simon (2022). "How Leibniz read Pascal's geometry: The practice of reading as a hybrid hermeneutic in the materiality of Leibniz's *Pascaliana*". *Nuncius*, forthcoming.
- Dumas Primbault, Simon, Paul-Arthur Tortosa, and Martin Vailly (eds.). (2021a). "Lieux et milieux de savoirs : pour une écologie des pratiques savantes". Special Issue of *Cahiers François Viète* III(10). [https://cfv.univ-nantes.fr/medias/fichier/vn-cahierscfv-serie-iii-10\\_1618391856047-pdf?ID\\_FICHE=1415775](https://cfv.univ-nantes.fr/medias/fichier/vn-cahierscfv-serie-iii-10_1618391856047-pdf?ID_FICHE=1415775)
- Dumas Primbault, Simon, Jérôme Baudry, Jean-François Bert, and Bayrem Kaabachi. (2021b). "« Des embranchements originaux ». Parcours de lecture et recherche exploratoire sur Gallica." Colloquium *Ethnographier les pratiques en régime numérique : enjeux et méthodes*, ENS Ulm, Paris. <http://infoscience.epfl.ch/record/288902?&ln=en>
- Ebli, Stefania, Michaël Defferrard, and Gard Spreemann. (2020). "Simplicial Neural Networks". <https://arxiv.org/abs/2010.03633>
- Edelsbrunner, Herbert, and John Harer. (2010). *Computational Topology: An Introduction*. Providence: American Mathematical Society.
- Franklin, Seb. (2015). *Control: Digitality as Cultural Logic*. Cambridge, MA: MIT Press.
- Galinon-Méléneç, Béatrice (ed.). (2011). *L'Homme-trace. Perspectives anthropologiques des traces contemporaines*. Paris: CNRS Éditions.
- Galinon-Méléneç, Béatrice, and Sami Zlitni (eds.). (2013). *Traces numériques. De la production à l'interprétation*. Paris: CNRS Éditions.
- Galinon-Méléneç, Béatrice, Sami Zlitni, and Fabien Liénard (eds.). (2015). *L'Homme-trace. Inscriptions corporelles et techniques*. Paris: CNRS Éditions. <https://doi.org/10.4000/books.editions-cnrs.25549>.
- Galloway, Alexander R. (2012). *The Interface Effect*. Cambridge, UK: Polity Press.
- Geertz, Clifford. (1973). *The Interpretation of Culture*. New York: Basic Books.
- Ghitalla, Franck (ed.). (2002). "La navigation". Special issue of *Les cahiers du numérique* 3(3).
- Ghitalla, Franck, Dominique Boullier, Pergia Gkouskou-Giannakou, Laurence Le Douarin, and Aurélie Neuau (eds.). (2013). *L'outre-lecture. Manipuler, (s')approprié, interpréter le Web*. Paris: Bibliothèque Centre Pompidou. <https://doi.org/10.4000/books.bibpompidou.463>
- Ghitalla, Franck. (2021). *Qu'est-ce que la cartographie du web ? Expéditions scientifiques dans l'univers des données numériques et des réseaux*. Marseille: OpenEdition Press. <https://doi.org/10.4000/books.oep.15358>
- Ginsparg, Paul. (2011). "It was twenty years ago today...". <https://arxiv.org/abs/1108.2700v2>
- Ginsparg, Paul (2017). "Preprint Déjà Vu: an FAQ". <https://arxiv.org/abs/1706.04188v2>
- Ginsparg, Paul (2021). "Lessons from arXiv's 30 years of information sharing". *Nature Reviews Physics* 3, 602-603. <https://doi.org/10.1038/s42254-021-00360-z>
- Ginzburg, Carlo. (1979). "Spie. Radici di un paradigma indiziario". In: Aldo Gargani (ed.). *Crisi della ragione. Nuovi modelli nel rapporto tra sapere e attività umane*. Turin: Einaudi.
- Gitelman, Lisa. (2014). *Paper Knowledge: Toward a Media History of Documents*. Durham: Duke University Press.

- Given, Lisa M., and Rebekah Wilson. (2018). "Information technology and the humanities scholar: Documenting digital research practices". *Journal of the Association for Information Science and Technology* 69(6), 807-819. <https://doi.org/10.1002/asi.24008>
- Gläser, Jochen, Wolfgang Glänzel, and Andrea Scharnhorst. (2017). "Same data — different results? Towards a comparative approach to the identification of thematic structures in sciences". *Scientometrics* 111, 979-979. <https://doi.org/10.1007/s11192-017-2296-z>
- Golumbia, David. (2009). *The Cultural Logic of Computation*. Cambridge, MA: Harvard University Press.
- Goody, Jack. (1977). *The Domestication of the Savage Mind*. Cambridge, UK: Cambridge University Press.
- Hargittai, Eszter, and Christian Sandvig (eds.). (2015). *Digital Research Confidential: The Secrets of Studying Behavior Online*. Cambridge, MA: MIT Press.
- Hookway, Branden. (2014). *Interface*. Cambridge, MA: MIT Press.
- Huberman, Bernardo A. (2001). *The Laws of the Web: Patterns in the Ecology of Information*. Cambridge, MA: MIT Press.
- Kittler, Friedrich A. (1990). *Discourse Networks 1800/1990*. Stanford: Stanford University Press.
- Knorr Cetina, Karin. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.
- Koopmann, Tobias, Alexander Dallmann, Lena Hettinger, Thomas Niebler, and Andreas Hotho. (2019). "On the right track! Analysing and Predicting Navigation Success in Wikipedia". In: *Proceedings of the 30th ACM Conference on Hypertext and Social Media (HT '19)*. New York: Association for Computing Machinery, 143-152. <https://doi.org/10.1145/3342220.3343650>
- Kortemeyer, Gerd, and Stefan Droschler. (2021). "A user-transaction-based recommendation strategy for an educational digital library". *International Journal on Digital Libraries* 22(2), 147-157. <https://doi.org/10.1007/s00799-021-00298-8>
- Krajewski, Markus. (2018). *The Server: A Media History from the Present to the Baroque*. New Haven: Yale University Press.
- Lamprecht, Daniel, Florian Geigl, and Tomas Karas. (2015). "Improving recommender system navigability through diversification: A case study of IMDb". *i-KNOW '15*. <http://dx.doi.org/10.1145/2809563.2809603>
- Lamprecht, Daniel, Kristina Lerman, Denis Helic & Markus Strohmaier. (2017). "How the structure of Wikipedia articles influences user navigation". *New Review of Hypermedia and Multimedia* 23(1), 29-50. <http://dx.doi.org/10.1080/13614568.2016.1179798>
- Lanclos, Donna, and Andrew D. Asher. (2016). "'Ethnographish': The State of the Ethnography in Libraries". *Weave: Journal of Library User Experience* 1(5) <https://doi.org/10.3998/weave.12535642.0001.503>
- Larivière, Vincent, Cassidy R. Sugimoto, Benoit Macaluso, Staša Milojević, Blaise Cronin, and Mike Thelwall. (2014). "arXiv E-prints and the journal of record: An analysis of roles and relationships". *Journal of the Association for Information Science and Technology* 65(6), 1157-1169. <https://doi.org/10.1002/asi.23044>
- Latour, Bruno. (1986). "Visualization and Cognition: Drawing Things Together". *Knowledge and Society - Studies in the Sociology of Culture Past and Present*, 6(0), 1-40. <http://www.bruno-latour.fr/sites/default/files/21-DRAWING-THINGS-TOGETHER-GB.pdf>
- Le Deuff, Olivier. (2016). "Humanités digitales et bibliothèques. Des milieux de savoir en quête de nouvelles compétences". *Balisages* 3. <https://publications-prairial.fr/balisages/index.php?id=133>
- Le Marec, Joëlle, and François Mairesse. (2017). *Enquête sur les pratiques savantes ordinaires. Collectionnisme numérique et environnements matériels*. Lormont: Le bord de l'eau.
- Leibon, Gregory, and Daniel N. Rockmore (2013). "Orienteering in Knowledge Spaces: The Hyperbolic Geometry of Wikipedia Mathematics". *PLoS ONE* 8(7), e67508. <https://doi.org/10.1371/journal.pone.0067508>
- Lenoir, Timothy (ed.). (1997). *Inscribing Science: Scientific Texts and the Materiality of Communication*. Stanford: Stanford University Press.
- Lury, Celia, Luciana Parisi, and Tiziana Terranova (eds.). (2012). "Topologies of Culture." Special issue of *Technology, Culture & Society*, 29(4-5). <https://doi.org/10.1177/0263276412454552>
- McKiernan, Gerry. (2000). "arXiv.org: the Los Alamos National Laboratory e-print server". *International Journal on Grey Literature* 1(3), 127-138. <https://doi.org/10.1108/14666180010345564>
- Merzeau, Louise (ed.). (2010a). "Présence numérique : de la gestion d'une identité à l'exercice d'une liberté". Special Issue of *Documentaliste-Sciences de l'information* 47(1). <https://doi.org/10.3917/docsi.471.0030>
- Merzeau, Louise. (2010b). "L'intelligence de l'utilisateur". Séminaire INRIA. L'utilisateur numérique, ADBS éditions, 9-37. <https://halshs.archives-ouvertes.fr/halshs-00526527>
- Mounier, Pierre. (2010). "Le livre et les trois dimensions du cyberspace". In: Marin Dacos (ed.). *Read/Write Book. Le livre inscriptible*. Marseille: OpenEdition. <https://doi.org/10.4000/books.oep.128>
- Mounier, Pierre. (2020). "Les plateformes de diffusion numérique, un enjeu politique". In: Étienne Anheim and Livia Foraison (eds.). *L'édition en sciences humaines et sociales. Enjeux et défis*. Paris: Éditions de l'École des hautes études en sciences sociales, 81-88. <https://doi.org/10.4000/books.editioneshess.28843>
- Nicholas, David, and David Clark. (2012). "Reading" in the Digital Environment. *Learned Publishing* 25(2), 93-98.
- Nouvellet, Adrien, Valérie Beaudouin, Florence D'Alché-Buc, Christophe Prieur, and François Roueff. (2017). "Analyse des traces d'usage de Gallica: Une étude à partir des logs de connexions au site Gallica". [Research Report] Télécom ParisTech; BnF. <https://hal.archives-ouvertes.fr/hal-01709264>

- O'Day, Vicki L. (1993). "Orienteering in an Information Landscape: How Information Seekers Get From Here to There". *INTERCHI '93*, 438-445. <https://doi.org/10.1145/169059.169365>
- Patania, Alice, Giovanni Petri, and Francesco Vaccarino. (2017). "The shape of collaborations". *EPJ Data Sci.* 6(18). <https://doi.org/10.1140/epjds/s13688-017-0114-8>
- Pignier, Nicole. (2009). "Sémiotique du webdesign : quand la pratique appelle une sémiotique ouverte". *Communication & langages* 159(1), 91-110. <https://doi.org/10.4074/S0336150009001082>
- Piper, Andrew. (2013). "Reading's Refrain: From Bibliography to Topology". *ELH* 80(2), 373-399. <https://doi.org/10.1353/elh.2013.0022>
- Pohl, Stefan. (2006). "Using Access Data for Paper Recommendations on ArXiv.org". MA Thesis, Technische Universität Darmstadt. <https://arxiv.org/abs/0704.2963>
- Priestner, Andy and Borg, Matt (eds.). (2016). *User Experience in Libraries: Applying Ethnography and Human-Centered Design*. New York: Routledge.
- Rieder, Bernhard. (2020). *Engines of Order: A Mechanology of Algorithmic Techniques*. Amsterdam: Amsterdam University Press.
- Rodighiero, Dario. (2021). *Mapping Affinities: Democratizing Data Visualization*. Geneva: Metis Presses.
- Roustan, Mélanie (ed.). (2016). *La recherche dans les institutions patrimoniales. Sources matérielles et ressources numériques*. Villeurbanne: Presses de l'ENSIB. <https://doi.org/10.4000/books.pressesensib.5682>
- Sack, Warren. (2019). *The Software Arts*. Cambridge, MA: MIT Press.
- Salonius, Pippa, and Michael D. J. Bintley (eds.). (2021). *The Surrounding Forest: Structuring Knowledge with Trees at the Time of the European Middle Ages*, Birkbeck College, forthcoming.
- Sarrazin, Véronique, Patrice Marcilloux, Bénédicte Grailles et Valérie Neveu (eds.). (2019). *Classer les archives et les bibliothèques. Mise en ordre et raisons classificatoires*. Rennes: Presses Universitaires de Rennes. <https://doi.org/10.4000/books.pur.88593>
- Schafer, Valérie (dir.). (2018). *Temps et temporalités du Web*. Nanterre: Presses universitaires de Paris Nanterre. <https://doi.org/10.4000/books.pupo.5988>
- Sen, Rituparna, and Mark H. Hansen. (2003). "Predicting Web Users' Next Access Based on Log Data". *Journal of Computational and Graphical Statistics* 12(1), 143-155. <https://doi.org/10.1198/1061860031275>
- Shiffrin, Richard M., and Katy Börner. (2004). "Mapping Knowledge Domains". *Proceedings of the National Academy of Sciences of the United States of America* 101, 5183-5185. <https://doi.org/10.1073/pnas.0307852100>
- Singla, Adish, Ryen W. White, and Jeff Huang. (2010). "Studying trailfinding algorithms for enhanced web search". *SIGIR'10*, 443-450. <https://doi.org/10.1145/1835449.1835524>
- Snee, Helene, Christine Hine, Yvette Morey, Steven Roberts, and Hayley Watson (eds.). (2016). *Digital Methods for Social Science: An Interdisciplinary Guide to Research Innovation*. Houndmills: Palgrave MacMillan.
- Souchier, Emmanuël, Yves Jeanneret, and Joëlle Le Marec (eds.). (2003). *Lire, écrire, réécrire. Objets, signes et pratiques des médias informatisés*. Paris: Bibliothèque publique d'information. <https://doi.org/10.4000/books.bibpompidou.394>
- Szoniecky, Samuel. (2018). *Ecosystems Knowledge: Modeling and Analysis Method for Information and Communication*. Hoboken: Wiley.
- Teevan, Jaime, Christine Alvarado, Mark S. Ackerman, and David R. Karger. (2004). "The perfect search engine is not enough: a study of orienteering behavior in directed search". *CHI 2004*, 415-422. <https://doi.org/10.1145/985692.985745>
- Theviot, Anaïs. (2021). "Confinement et entretien à distance : quels enjeux méthodologiques ?". *Terminal* 129. <https://doi.org/10.4000/terminal.7193>
- Tkacz, Nathaniel. (2015). *Wikipedia and the Politics of Openness*. Chicago: The University of Chicago Press.
- Tresch, John (ed.). (2014). *Bibliotechnica. Humanist Practice in Digital Times*. Venice: Fondazione Cini.
- Trace, Ciaran B., and Unmil P. Karadkar. (2017). "Information management in the humanities: Scholarly processes, tools, and the construction of personal collections". *Journal of the Association for Information Science and Technology* 68(2), 491-507. <https://doi.org/10.1002/asi.23678>
- Venturini, Tommaso, Dominique Cardon, and Jean-Philippe Cointet. (2014). "Méthodes digitales. Approches quali/quantitative des données numériques". Special Issue of *Réseaux* 6(188), 9-21. <https://doi.org/10.3917/res.188.0009>
- Weiland, Steven. (2021). *The Scholarly Workflow in the Digital Age*. Sullivan's Island: Against the grain. <https://doi.org/10.3998/mpub.12119260>
- White, Ryen W., and Jeff Huang. (2010). "Assessing the Scenic Route: Measuring the Value of Search Trails in Web Logs". *SIGIR'10*, 587-594. <https://doi.org/10.1145/1835449.1835548>
- White, Ryen W., and Adish Singla. (2011). "Finding our way on the web: exploring the role of waypoints in search interaction". *WWW 2011*, 147-148. <https://doi.org/10.1145/1963192.1963267>
- Zoellner, Kate, Samantha Hines, Teresa Keenan, and Sue Samson. (2015). "Faculty Research and Publication Practices". *portal: Libraries and the Academy* 15(1), 111-131. <https://doi.org/10.1353/pla.2015.0009>