UNFOLDING FRICTIONS IN DATABASE PROJECTS

Florian Jaton, Dominique Vinck

S.A.C. | « Revue d'anthropologie des connaissances »

2016/4 Vol. 11 | pages a à m

UNFOLDING FRICTIONS IN DATABASE PROJECTS

FLORIAN JATON
DOMINIQUE VINCK

The Dead Sea Scrolls on the doorstep of a new database

The Dead Sea Scrolls were discovered throughout the 1950s in caves of the area of Qumran, West Bank. Today they constitute a corpus of about 950 manuscripts, probably written between the end of the third century BCE and the middle of the first century CE. Their study gives us precious elements about the spiritual shifts of this period of Judaic transformations and early Christianity. The quality of the manuscripts varies considerably, going from well-preserved scrolls to tiny fragments consumed by fire and *fungi*. The collective, laborious and controversial decipherment of the manuscripts has driven digitization projects from as early as the 1960s. Today, the many digital systems available to explore and study the corpus include infrared luminescence technologies that reveal characters invisible to the naked eye, tools for textual segmentation and vocalization, integrated dictionaries that can access the "original" scanned texts in high-definition, and reconstitution tools for lost passages based on formal handwritten characteristics (outlines of characters, spacing deviation between words and characters, etc.). Google has also recently launched a re-digitization project of the corpus in ultra-high-definition in collaboration with the Israeli government. Many tools and systems for the analysis of this corpus – made famous through the popular novel *The Da Vinci Code* – coexist then in a quite competitive environment.

The project of David Hamidovic – a historian of antiquity – is in line with recent attempts to exploit the pixel properties of the scanned manuscripts. Yet, contrary to what has been already proposed, his ambition is to learn more about the important yet mysterious people who wrote these manuscripts: their *scribes*. His question is the following: what are the *editorial projects* inscribed within the manuscripts? The question is a legitimate one since at that time, scribes did not copy (monasteries are later historical formations): they tell,

add, rearrange and modify according to different sensibilities and affects. If they did not radically transform the final messages of canonical texts, scribes often added or changed some elements of the stories. And since they had diverse sensibilities, backgrounds, and references, these scribes can be considered as expressers – and, to a certain extent, actors – of the spiritual shifts of this historical period. By collecting and comparing semantic and formal *traces* left behind by scribes (words, turns of phrase, curves of the characters, spacing between words), it surely is possible to build *profiles* of scribes with diverse interests and editorial projects. And today, this could be done with statistical rigor via the creation of a database with structuration and content that could support the definition and training of more or less supervised machine-learning algorithms.

Yet, from this point, the questions that arise *immediately* concern the protocol required in order to shape this database. Since the manuscripts are quite renowned, it is not difficult to raise the interest of computer science laboratories specialized in the automated processing of hand-written characters. But the development of machine-learning models necessitates heavy preparatory work: how could a historian of antiquity foresee the irreversible consequences of the computational propositions he is being proposed? How can he even understand these propositions? And similarly, how could computer scientists foresee the potential dangers of their database model for the practice of history? How do they start the construction of the preparatory database of digital characters without making important mistakes that are too expansive to correct for both parties? In sum, how do they start such an interdisciplinary project in the best possible way for *both* the history of antiquity *and* two-dimensional digital signal processing?

Many database projects are under way in the human and social sciences (HSS). From small PhD projects to big national infrastructures, an increasing amount of researchers in HSS endeavor to collect, select, and organize digital data, with the help of computer scientists, in order to make them query*able*, share*able*, and eventually augment*able* (Magnien & Vinck, to be published). Moreover, after a two-year research seminar where colleagues were invited to discuss problems related to their database projects, it appeared to us that most of these projects were sources of controversies and disputes. Science and human *frictions* (Edwards *et al.*, 2011) seem to emanate from these projects, notably questions about the status of HSS in the academic landscape, the adequacy of digital tools for research, and the credibility and utility of HSS for the collective world. Often involuntarily, humanists and social scientists engaged in such projects also carry visions for their disciplines and become controversial prescribers of good practices and technical solutions.[1] During these work situations, real

1 Cf. the review of *Understanding Research Infrastructures in the Social Sciences* (Kleiner *et al.,* 2013) by Vinck (2013). The studies on *Big Science* and its particle accelerators (Simoulin, 2007; Boisot et al.,

trials (Boltanski and Thévenot, 2001) take place with results that tend to set up irreversibility consequences.

Yet, even though trials and controversies are now common sociological starting points (Akrich *et al*., 2006; Boltanski & Thévenot, 1991),[2] very few *studies* have inquired into these hectic projects. Despite their noisy aspects and relatively easy access – often the department next door, or even one's own lab! – that make them ideal environments for sociological explorations, very few serious attempts have been made to account for these projects. Some researchers have published their reflections but they mainly constitute prescriptive positions about scientific uses of digital technologies.[3] The recent popularity of social studies of digital databases (cf. below) makes the story even more surprising: even though heuristic controversial processes engaging digital technologies and the production of knowledge in HSS are happening next door, most inquiries seem to focus on other disciplines.

By diving into the shaping of digital databases in HSS, this special issue tries to better understand what is going on in and through these conflicting situations. By focusing on *frictional moments* (Edwards *et al*., 2011), each contribution participates in documenting the *mode of the existence* (Latour, 2012) of digital databases as well as the dynamics of *knowledge infrastructures* (Karasti *et al*., 2016a, 2016b) for HSS. As we shall see, these contributions reveal new injunctions, competences and articulations that surely deserve greater attention.

SOCIAL STUDIES OF DATABASES

Social studies of databases are not in their infancy, even though they do not constitute a unified sociological tradition. According to us, four major research directions can be listed.

The first research direction is sociohistorical and is the result of recent efforts made by the *Annals of the History of Computing* (Grad & Bergin, 2009 ; Grad, 2012, 2013). Quite ignored in the French-speaking academic landscape, these works have brought into light the sociogenesis of the term "database" (Haigh, 2009) as well as its attachments to administrative, military and managerial communities (Bergin & Haigh, 2009; Haigh, 2011; Grier, 2012; Wade & Chamberlin, 2012; Haderle & Saracco, 2013). If digital databases often appear as confident extensions of digital data collections, these sociohistorical inquiries have documented their fragile beginnings and their progressive

2011), radio telescopes, nanotechnology platforms (Merz & Biniok, 2010; Hubert, 2013) and on genomics (Peerbaye, 2004), scientific cooperation networks (Vinck, 1999, 2012) and natural histories (Bowker & Star, 1999; Bowker, 2005) have not inspired similar studies on HSS.

2 Lemieux (2007) distinguishes between two ways to consider controversies: a "classical way" that considers them as revealers of previous sociological positions and a "science studies way" that consider them as performative processes that establish new entities.

3 See for example Currie (2012), Heftberger (2012) and Kirschenbaum (2007).

associations with technologies that today appear ubiquitous (relational models, SQL language, theoretical distinction between "data" and "views," etc.).

The second research direction analyzes the *effects* digital databases have *on* society. Drawing upon surveillance studies, they tend to denounce the subjection power of databases (Lyon, 2003) and their ability to produce restrictive categories (Porter, 1995; Atten, 2013). If this reading of databases tends to decry the asymmetry between those who design databases (private companies, administrations) and those who endure them, actors are not always considered as passive agents and do sometimes manage to reconfigure them through unexpected interactions and uses (Flichy, 2013). This emphasis on the effects of databases also encapsulates symbolic and cultural interpretations (Manovich, 1999, 2012) where digital databases are considered as new media reflecting today's society. This interpretation in terms of *mediology* puts the emphasis on the new potentials for action inscribe within digital databases.

A third research direction – mainly initiated by science and technology studies (STS) – puts the emphasis on the shaping processes of digital databases. Yates (2005), Segresting (2004), Gardey (2008) and Vinck & Penz (2008) have shown that the initial desires that trigger database projects are radically transformed once confronted with construction issues, hence creating new arrangements. Similarly, other studies have explored the shifts between initial goals and effective uses. Ugettho (2013) shows for example that intertwined co-construction mechanisms easily widen the gap between imagined and effective uses of digital databases. Discussions, disagreements and many unexpected events transform the architectures and functionalities of databases that sometimes end up expressing something very different from their designers' initial plans. Moreover, Dagiral and Peerbaye (2013) show how a digital database initially designed to share information is confronted with equivocal interpretations of this information. This leads to struggles among users and designers to establish a "standardized" interpretation of the shared information. According to this research direction, it is then difficult to infer anything from the *effects* of databases; one should rather document the performativity of their collective construction processes.

A fourth direction of analysis – close to the third one – considers databases as part of broader *categorization and classification processes*. Porter (1995) for example inquires into the social, political and economic dynamics that shaped the categories of an American administrative database. He shows how private companies, competing administrations, federal agencies, social scientists, programmers and computer scientists more or less successfully influenced the definition of the categories to their advantages. In the same vein, the work of Susan Leigh Star and Goeffrey C. Bowker (2000) shows that classification practices required by databases go along with *invisibilization processes*; the collection and structuration of data also implies more or less harmful exclusions and muting (Star, 1989; Bowker, 2000). Using the example of a digital repository for cycle lanes, Denis and Pontille (2013) further show how collection practices

imply binding choices as well as specific competences. In their work on research infrastructures for biodiversity, Granjou *et al.* (2014) describe how the structuration of contents requires new collaborations and management systems that tend to transform taxonomists into "data providers." In the same vein, Dagiral and Peerbay (2016) show how the feeding, maintenance and use of a research database on rare diseases include a surprising duality: for the actors involved, this *knowledge infrastructure* (Karasti *et al.*, 2016a, 2016b) is stable and representative of rare diseases *but also and at the same time* problematic and potentially useful for further (re)negotiations. The invisibilization practices that go along with digital databases seem then to be coupled with systematic infrastructural inversion practices (Bowker, 1994) that are potential resources for the actors in situation, and not only researchers in social sciences.

As different and interesting as they are, these studies have in common that they largely ignore the relationships between digital databases and HSS and the mutations they may provoke. Indeed, notably through the rise of digital humanities (Magnien & Vinck, to be published), more and more funding for innovative projects that gather practitioners of HSS – historians, philosophers, psychologists, sociologists – and of computer science and technologies (CST) – data scientists, GUI designers, programmers, system architects – are allocated; what happens when these actors work together in order to design a database? How do (dis)agreements and compromises arise? What paths do collective practices of data structuration take? As we shall see in this special issue, compromises are instituted, attempts are made to put arrangements in place and new entities appear. Through the constitution of digital databases in HSS, a specific research environment is expressed possessing an ecology that may deserve greater consideration.

DATABASES IN THE SCIENCES AND HUMANITIES

For sure, databases are important entities for scientists: it is difficult indeed to imagine the practice of science without any analogical or digital device for data storage and retrieval (Bowker, 2008). According to Latour (1989), scientific activity necessitates the stabilization of resources and constants in order to produce traces and inscriptions; it is only through laborious compilations of documents that more or less controversial *chains of reference* (Latour, 2012) can be put in place. But traces, alignments, synthetizations and compilations also imply storage, exchanges and standards: the extension of a knowledge network requires *equipment* (Vinck, 1999, 2011) in order to make traces of events commensurable. Even before the rise of digital technologies, the constitution of data collections that could support specific requests and comparisons seems then coextensive to the scientific activity (Desrosières, 2008).

The dynamics related to the constitution of databases in the "hard" sciences are pretty well documented (Hilgarter, 1995, 2012; MacKenzie, 2003; Beaulieu, 2004; Bowker, 2000, 2005; Hine, 2006; Heaton & Proulx, 2012; Edwards, 2013; Meyer & Schroeder, 2015). Bruno Strasser (2011) shows, for example, that the rise of digital storage and compilation has partially transformed the field of genomics. Indeed, because of the growing emphasis on the collection, storage and sharing of data, this science that used to be presented as mainly *experimental* looks more and more like "natural sciences," older and often less valorized. Notably because of the rise of digital technologies and the setting up of enormous databases (*big data*), practitioners of genomics collect, sort, standardize and share data more than they organize "classical" analogical laboratory experiments. A different way to conceive and present genomics is shaped through the potentialities of digital databases and this, of course, causes resistances and transformations. But what about HSS? How do digital databases transform their practices? Despite the importance of the topic, very few studies have been conducted (Vinck, 2016).

Digital data collection in the human and social sciences

Databases are also important in HSS. From small individual projects to big institutional strategies requiring costly technological infrastructures,[4] databases in HSS have histories that would be, according to us, important to unfold. We think, for example, about the constitution of private archives during the 1940s in the United States for the study of public opinion, the development of quantification in HSS (notably in linguistics, history and sociology), the constitution of national statistic resources and thesauri, and the recent development of digital humanities and their relationships with the notion of *big data*; all these topics are ultimately linked with the design of databases. The structuration and sharing of data are also related to methodological issues about protocols, comparability, interoperability and quality control. Performative definitions of equivalence between objects, people, relationships, and events (Desrosières, 2008) required by quantitative analysis are also an issue: choices related to data structure within databases can produce very tangible effects, for example when HSS propositions are included in the design of policies. More inquiries into digital databases in HSS could help better understand these socio-technical processes that take part in the shaping of the collective world.

The constitution of databases in HSS also implies best practices and international norms – such as the *Data Documentation Initiative* (DDI) and the *International Standard Archival Description* (ISAD) – for the collection, preparation, harmonization and cleaning of data as well as the documentation of their

4 Some European examples among many, the "Consortiums Eurpéens d'Infrastructures de Recherche" (ERIC), the "Conseil Européen pour l'Archivage des Données en Sciences Sociales," the *International Federation of Data Organisations* (IFDO) and, in France, the infrastructure "Huma-Num" to support research in the humanities and social sciences.

conditions of production and indexation (Beltrame & Jungen, 2013). Yet, even though these practices are highly problematic, the success of databases in HSS seems also to rely upon their invisibilization[5] since the users-researchers do not necessarily want to bother with conception, maintenance, or security issues. To a certain extent, the utility of databases in HSS seems to be also a function of their ability to be put into black boxes.

This tension between obvious issues and necessary blinders works to produce misunderstanding with CST specialists concerning both the conception and the utilization of databases in HSS. Whatever one might say, in-depth collective discussions upstream of the construction of databases remain badly equipped. The contracted computer scientist is often tasked with choosing the database management system, even though it may quickly create irreversible consequences or may not guarantee any sustainability.

Some issues related to digital data collections in HSS

Pressures for results certainly contribute to the under-exploration of issues and tensions that arise during the conception and management of databases in HSS; for obvious reasons, researchers tend to privilege demos and publications in ranked journals over in-depth reflections about their own practices. Yet we do think that these projects are perfect situations for learning more about our disciplines: by documenting what is going on during these eventful work situations, one may produce refreshing knowledge about research practices in HSS when dealing with digital data collections. An unpublished preliminary inquiry conducted by Pierre-Nicolas Oberhauser has stressed some of these issues that will be further explored by the contributions of this special issue.

Issues related to the duplication and sharing of digital data

When data were confined in binders and cardboard files, researchers in HSS had to produce very tangible efforts in order to share them with their colleagues. Once digitally translated, data seem more easily duplicable and the efforts required to share them seem to diminish: other entities – operating systems, data transmission protocols, systems architects – are in fact taking over, hence diminishing the costs of circulation, at least from the point of view of the researchers. Data then seem more sharable and better suited for distributed projects.

These seemingly intrinsic sharing capabilities of digital data do not only intervene *downstream* of their production: they also intervene *upstream*, when they are shaped with the help of digital tools. Hence the apparition of other issues notably related to their property: who owns digital data? Those who own

5 See the special issue on this topic in the *Revue d'anthropologie des connaissances* – Denis and Pontille (2012), especially the paper of Millerand (2012).

the servers that store them? The managers of the research projects? Those who equipped them in order to make them able to circulate within communication networks? This further indicates issues related to the equipment of data – data are never sharable *by themselves* but always *by others*: who are those who work in order to make them suitable for digital networks? Are they not often subordinated employees doing repetitive tasks that may deserve greater consideration?

Issues related to the collectivization of scientific work

The collectivization of scientific work suggested by the sharing capabilities of digital data further creates organizational issues. The first one is related to a *bootstrapping problem* (Bowker *et al.*, 2010). Indeed, in order to be considered as trustworthy, one set of data on a specific topic often needs to be crosschecked with another set of data produced by other researchers on the *same* topic. From this point, the collectivization of scientific work induced by digitization further induces the definition of valuable *topics* capable of bridging diverse preoccupations. These "boundary-objects" are of course problematic precisely because of their power of attraction: how are they constituted and how can we eventually make them shift (Star & Griesemer, 1989; Star, 2010)?

A second issue related to the collectivization of scientific work is related to the standardization of data and the definition of formatting rules: how are they produced? And how do researchers in HSS enact them?

Quality issues

Once produced – that is shaped, organized and equipped so that they can be shared and queried – data are often submitted to epistemological proofs that test their robustness (what was the data collection protocol? what references were used for the conceptualization of the research? etc.). But equally important are the technical proofs that test the ability of data to suit the practical and changing needs of researchers (are the data in a standardized format? could they be merged with another dataset without compatibility problems? etc.). Robustness of the data and flexibility of the collection: a subtle articulation between these two attributes seems to define a *good* digital data collection for HSS. Yet are these criteria always considered evenly?

FRICTIONS THAT "MAKE DO"

More than analyzing the *effects* digital databases have *on* HSS, this special issue tries to document the transformative relationships that pop up during the construction of digital databases: what do these processes make us do? Over the active voice of control and the passive voice of domination, this special issue privileges the middle voice of the *making do* (Latour, 2000). What do the sparks

of *trials* (Akrich, Callon, & Latour, 2006 ; Boltanski & Thévenot, 1991) and *frictions* (Edwards *et al.*, 2011) draw when digital databases are being designed for HSS?

Each of the contributions answers this question in its own way. For Gilles Bastin and Jean-Marc Francony, who try to transform *inscriptions* stored in the Web servers of LinkedIn into *data* capable of being sociologically queried, the unfolding of frictions reveal a need of *visibilization*: the transformation process that makes Web inscriptions become sociological data "wakes up" unexpected actors whose views need to be taken into account. The design of their database is then coupled with inquiries into these contradictory views in order to become capable of composing with them. In short, rather than muting discordant views and interests, the construction of their database requires their expressions and considerations.

Sophie Duchesne and Mathieu Brugidou analyze the constitution of a database for French-speaking qualitative inquiries. By exploring the numerous frictions of this project, they make appear a continuous need of *simulation*: throughout its progress, the project never stops asking its participants to foresee—and eventually confront – what it could later enable, the difficulty being precisely the variability of these projections through time. The more entities are included into the project – funding, people, institutions—the more difficult it becomes for those involved to confront their visions and even more to conciliate them.

Finally, by finely exploring the interactions between actors conducting a database project in psychology, Pierre-Nicolas Oberhauser documents the competences that need to be constructed – almost on the fly! – in order to provoke relationships between domains of expertise. The frictions of a database project are therefore not given: the actors need to develop situational competences in order to eventually make them appear.

These three contributions unfold the frictions of database projects in HSS in quite different ways: are they commensurable enough to support a broader proposition that would merge them? It is, of course, up to the readers to decide. Yet, at least according to us, these contributions do not only explore situations that were previously poorly documented; they also help digital databases to become more *familiar* by attributing to them *composition*, *simulation* and *competences* desires. These conceptual supports may engage new inquiries but they may also help the pursuit of new interdisciplinary database projects.

Acknowledgments

The publication of this special report was supported by the Fonds des publications of the University of Lausanne (Unil). We would like to thank the participants of the LaDHUL (Laboratory of Digital Humanities of the University of Lausanne) workshop on databases held in October 2015 as well as the institutions that supported it, in particular the Institute of Social Sciences of the University of Lausanne. We are also very grateful to Marc Barbier who followed up this special issue with great care on behalf of the RAC Editorial Board.

REFERENCES

Akrich, M., Callon, M., & Latour B. (2006). *Sociologie de la traduction : textes fondateurs*. Paris : Presses de l'École des Mines.

Atten, M. (2013). Ce que les bases de données font à la vie privée. L'émergence d'un problème public dans l'Amérique des années 1960. *Réseaux*, *178-179*, 23-53. doi:10.3917/res.178.0021

Beaulieu, A. (2004). From brainbank to database: the informational turn in the study of the brain, *Studies in History and Philosophy of Science. Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *35*(2), 367-390.

Beltrame, N.T. & Jungen, C. (2013). Cataloguer, indexer, encoder. Ou comment les données prennent vie, *Revue d'anthropologie des connaissances*, *7*(4), 747-759. doi:10.3917/rac.021.0747

Bergin, T.J. & Haigh, T. (2009). The Commercialization of Database Management Systems, 1969-1983. *IEEE Annals of the History of Computing*, *31*(4), 26-41.

Boisot, M., Nordberg, M., Yami, S., & Nicquevert, B. (2011). *Collisions and collaboration. The organization of learning in the Atlas experiment at the LHC.* Oxford: Oxford University Press.

Boltanski, L. & Thévenot, L. (1991). *De la justification : les économies de la grandeur*. Paris : Gallimard.

Bowker, G.C. (1994). Information Mythology and Infrastructure, in L. Bud-Friedman (ed.), *Information Acumen: the Understanding and Use of Knowledge in Modern Business*, London: Routledge, 231-247.

Bowker, G.C. (2000). Biodiversity Datadiversity, *Social Studies of Science*, *30*(5), 643-683.

Bowker, G.C. (2005). *Memory practices in the sciences*. Cambridge, Mass., London: MIT Press.

Bowker, G.C. & Star, S.L. (2000). *Sorting things out. Classification and its consequences*. Cambridge, Mass.: MIT Press.

Bowker, G.C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment, in J. Hunsinger, L. Klastrup, & M. Allen (eds.), *International Handbook of Internet Research* (pp. 97-117). Dordrecht: Springer.

Currie, M. (2012). The Feminist Critique: Mapping Controversy in Wikipedia, in D. Berry (ed.), *Understanding Digital Humanities* (pp. 224-248). Houndmills: Palgrave Macmillan.

Dagiral, E. & Peerbaye, A. (2013). Voir pour savoir. Concevoir et partager des « vues » à travers une base de données biomédicales, *Réseaux*, *178-179*, 162-196. doi:10.3917/res.178.0163.

Dagiral, E. & Peerbaye, A. (2016). Making Knowledge in Boundary Infrastructures: Inside and Beyond a Database for Rare Diseases, *Science & Technology Studies*, *29*(2), 44-61.

Denis, J. & Pontille, D. (2012). Travailleurs de l'écrit, matières de l'information, *Revue d'anthropologie des connaissances*, *6*(1), 1-20. doi:10.3917/rac.015.0001

Denis, J. & Pontille, D. (2013). Une infrastructure élusive : aménagements cyclables et troubles de la description dans OpenStreetMap, *Réseaux*, *178*, 91-125.

Desrosières, A. (2008). *Pour une sociologie historique de la quantification*. Paris : La Découverte.

Edwards, P.N., Mayernik, M.S., Batcheller, A.L., Bowker, G.C., Borgman, C.L. (2011). Science Friction: Data, Metadata, and Collaboration. *Social Studies of Science*, *41*(5), 667-690. doi:10.1177/0306312711413314

Edwards, P.N. (2013). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*, Cambridge, Mass.: The MIT Press.

Flichy, P. (2013). Rendre visible l'information. Une analyse sociotechnique du traitement des données. *Réseaux*, *178-179*, 55-89.

Flichy, P. & Parasie, S. (2013). Sociologie des bases de données : présentation, *Réseaux*, *178-179*, 9-19. doi:10.3917/res.178.0009

Gardey, D. (2008). *Écrire, calculer, classer. Comment une révolution de papier a transformé les sociétés contemporaines (1800-1940)*. Paris : La Découverte.

Grad, B. (2012). Relational Database Management Systems: The Formative Years. *IEEE Annals of the History of Computing*, *34*(4), 7-8.

Grad, B. (2013). Relational Database Management Systems: The Business Explosion. *IEEE Annals of the History of Computing*, *35*(2), 8-9.

Grad B. & Bergin, T.J. (2009). History of Database Management Systems. *IEEE Annals of the History of Computing*, *31*(4), 3-5.

Granjou, C., Mauz, I., Barbier, M., & Breucker, P. (2014). Making taxonomy environmentally relevant. Insights from an all taxa biodiversity inventory. *Environmental Science & Policy*, *38*, 254-262.

Grier, D.A. (2012). The Relational Database and the Concept of the Information System. *IEEE Annals of the History of Computing*, *34*(4), 9-17.

Haderle, D.J. & Saracco, C.M. (2013). The History and Growth of IBM's DB2. *IEEE Annals of the History of Computing*, *35*(2), 54-66.

Haigh, T. (2009). How Data Got Its Base: Information Storage Software in the 1950s and 1960s. *IEEE Annals of the History of Computing*, *31*(4), 6-25.

Haigh, T. (2011). Charles W. Bachman: Database Software Pioneer. *IEEE Annals of the History of Computing*, *33*(4), 70-80.

Heaton, L. & Proulx, S. (2012). La construction locale d'une base transnationale de données en botanique. Une mise en lumière du travail des « petites mains », *Revue d'anthropologie des connaissances*, *6*(1), 141-162. doi:10.3917/rac.015.0179.

Heftberger, A. (2012). Do Computers Dream of Cinema? Film Data for Computer Analysis and Visualisation, in D. Berry (ed.), *Understanding Digital Humanities* (pp. 210-223). Houndmills: Palgrave Macmillan.

Hilgartner, S. (1995). Biomolecular Databases: New Communication Regimes for Biology?, *Science Communication*, *17*(2), 240-263.

Hilgartner, S. (2012). Selective flows of knowledge in technoscientific interaction: information control in genome research, *The British Journal for the History of Science*, *45*(2), 267-280.

Hine, C. (2006). Databases as scientific instruments and their role in the ordering of scientific work. *Social Studies of Science*, *36*(2), 269-298.

Hubert, M. (2013). *Partager des expériences de laboratoire. La recherche à l'épreuve des réorganisations*. Paris : EAC.

Jaton, F. (2014). Approches sociologiques des bases de données, miméo, compte rendu de l'exposé de P.N. Oberhauser et D. Vinck lors du séminaire du LaDHUL du 31 mars 2014, 4-5.

Karasti, H., Millerand, F., Hine, C.M., & Bowker, G.C. (2016). Knowledge infrastructures: Part I. *Science & Technology Studies*, *29*(1), 2-10.

Karasti, H., Millerand, F., Hine, C.M., & Bowker, G.C. (2016). Knowledge infrastructures: Part II. *Science & Technology Studies*, *29*(2), 2-6.

Kirschenbaum, M.G. (2007). *The remaking of reading: Data mining and the digital humanities*. The National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation, Baltimore, MD.

Kleiner, B., Renschler, I., Wernli, B., Farago, P., & Joye, D. (eds.) (2013). *Understanding Research Infrastructures in the Social Sciences*. Zurich: Seismo.

Latour, B. (1989). *La science en action. Introduction à la sociologie des sciences*. Paris : La Découverte.

Latour, B. (2000). Facture/fractures. De la notion de réseau à celle d'attachement. In A. Micoud & M. Peroni (dir.). *Ce qui nous relie* (pp. 189-208). La Tour d'Aigues : Éditions de l'Aube.

Latour, B. (2012). *Enquête sur les modes d'existence : une anthropologie des Modernes*. Paris : La Découverte.

Lemieux, C. (2007). À quoi sert l'analyse des controverses ? *Mil neuf cent. Revue d'histoire intellectuelle*, *25*(1), 191-212.

Lyon, D. (ed.) (2003). *Surveillance as social sorting: privacy, risk, and digital discrimination*. London: Routledge.

MacKenzie, A. (2003). Bringing sequences to life: how bioinformatics corporealizes sequence date, *New Genetics and Society*, *22*(3), 315-332.

Magnien, Y. & Vinck, D. (à paraître). Des résistances à la numérisation des données : les aventures du Big data au CNRS. In I. Sainsaulieu & A. Saint-Martin (dir.). *Travail, sciences et technologies : les formes contemporaines de l'innovation*.

Manovich, L. (1999). Database as Symbolic Form, *Convergence, 5*(2), 80-99. doi:10.1177/135485659900500206.

Manovich, L. (2012). How to Compare One Million Images? In *Understanding Digital Humanities*, Palgrave Macmillan, 249-278.

Merz, M. & Biniok, P. (2010). How Technological Platforms Reconfigure Science-Industry Relations: The Case of Micro- and Nanotechnology, *Minerva*, *48*, 105-124.

Meyer, E.T. & Schroeder, R. (2015). *Knowledge Machines: Digital Transformations of the Sciences and Humanities*. Cambridge, Mass.: MIT Press.

Millerand, F. (2012). La science en réseau. Les gestionnaires d'information « invisibles » dans la production d'une base de données scientifiques, *Revue d'anthropologie des connaissances*, *6*(1), 163-190.

Parasie, S. (2013). Des machines à scandale. Élément pour une sociologie des bases de données, *Réseaux*, *178-179*, 127-161. doi:10.3917/res.178.0127

Peerbaye, A. (2004). *La construction de l'espace génomique en France : la place des dispositifs instrumentaux*. Thèse de doctorat, Cachan, École Normale Supérieure de Cachan.

Porter, T.M. (1996). *Trust in Numbers*. Reprint Edition. Princeton, N.J.: Princeton University Press.

Segrestin, D. (2004). *Les chantiers du manager*. Paris : Armand Colin.

Simoulin, V. (2007). Une communauté instrumentale divisée... et réunie par son instrument, *Revue d'anthropologie des connaissances*, *1*(2), 221-241.

Star, S.L. (1989). *Regions of the Mind: Brain Research and the Quest for Scientific Certainty*. Stanford: Stanford University Press.

Star, S.L. (2010). Ceci n'est pas un objet-frontière ! Réflexions sur l'origine d'un concept, *Revue d'anthropologie des connaissances*, *4*(1), 18-35. doi:10.3917/rac.009.0018

Star, S.L. & Griesemer, J. (1989), Institutionnal ecology, 'Translations' and Boundary objects: amateurs and professionals on Berkeley's museum of vertrebate zoology. *Social Studies of Science*, *19*(3), 387-420.

Star, S.L. & Strauss, A. (1999). Layers of silence, arenas of voice: the ecology of visible and invisible work, *Computer Supported Cooperative Work*, *8*, 9-30.

Strasser, B. (2011). The Experimenter's Museum: GenBank and the moral economies of biomedecine, *ISIS*, *102*(1), 60-96.

Ughetto, P. (2013). Utiliser une base de données en organisation. La recherche de l'instrument, *Réseaux, 178-179*, 197-222. doi:10.3917/res.178.0197

Vinck, D. (1999). Les objets intermédiaires dans les réseaux de coopération scientifique. Contribution à la prise en compte des objets dans les dynamiques sociales, *Revue française de sociologie*, *40*(2), 385-414.

Vinck, D. (2011). Taking intermediary objects and equipping work into account when studying engineering practices, *Engineering Studies*, *3*(1), 25-44.

Vinck, D. (2013). Pour une réflexion sur les infrastructures de recherche en sciences sociales, *Revue d'anthropologie des connaissances*, *7*(4), 993-1001.

Vinck, D. & Penz, B. (dir.) (2008). *L'équipement de l'organisation industrielle. Les ERP à l'usage*. Paris : Hermès.

Wade, B.W. & Chamberlin, D.D. (2012). IBM Relational Database Systems: The Early Years. *IEEE Annals of the History of Computing*, *34*(4), 38-48.

Yates, J.A. (2005). *Structuring the Information Age: Life Insurance and Technology in the Twentieth Century*. The Johns Hopkins University Press.

Florian JATON is PhD student in social study of science and technology at the University of Lausanne. He is part of the Institute of Social Sciences, the STS Lab and the Laboratory of Digital Humanities of the University of Lausanne (LaDHUL). He is currently doing a research stay at the University of California at Irvine, in the EVOKE laboratory. His thesis is an ethnographic account of a computer science laboratory specialized in digital signal processing. This laboratory study deals more specifically with referential repositories for algorithmic procedures, programming practices and the shaping of mathematical models.

Adresse	Université de Lausanne (UNIL) Institut des Sciences Sociales – LaDHUL Quartier Mouline - Géopolis CH-1015 Lausanne (Suisse)
Courriel	florian.jaton@unil.ch, fjaton@uci.edu

Dominique VINCK is full Professor at the University of Lausanne (UNIL) and teaches at the College of Humanities at the Swiss Federal Institute of Technology in Lausanne (EPFL). He is member of the Social Sciences Institute and STS Lab of the UNIL and director of the Laboratory of Digital Humanities of the University of Lausanne (LaDHUL). His research interests are in the sociology of science and innovation and the engineering of digital humanities and cultures. His publications include *Pratiques de l'interdisciplinarité* (PUG, 2000), *Everyday Engineering. An Ethnography of Design and Innovation* (MIT Press, 2003), *The Sociology of Scientific Work* (Edward Elgar, 2010), *L'équipement de l'organisation industrielle. Les ERP à l'usage* (Hermes, 2008), *Les nanotechnologies* (Le Cavalier Bleu, 2009), *Comment les acteurs s'arrangent avec l'incertitude* (EAC, 2009), *Les Masques de la convergence* (EAC, 2012), *Ingénieur aujourd'hui* (PPUR, 2015), *Sciences et technologies émergentes : pourquoi tant de promesses ?* (Hermann, 2015), *Humanités numériques : la culture face aux nouvelles technologies* (Le Cavalier Bleu, 2016).

Affiliation	Université de Lausanne (UNIL) Institut des Sciences Sociales – LaDHUL Quartier Mouline - Géopolis CH-1015 Lausanne (Suisse)
Courriel	dominique.vinck@unil.ch