# Archimorphosis

Brigitte Ouvry-Vial, Yannick Estève

# Archimorphosis / Brigitte Ouvry-Vial, Yannick Estève

**Published in 'Archives on fire', a special issue of journal *Reconstruction*, May 2016**

<1> Computers now offer opportunities to grasp, store, and explore huge corpora in a matter of weeks or months, and in a proportion ten years ago barely conceivable to scholars in the Humanities who would not have taken up the task of arranging such corpora with traditional archival and documentation tools, less committing themselves to a lifetime of work. Recent algorithms and hardware also allow computer scientists to develop new software tools based on machine learning and artificial intelligence, in order to emphasize inconspicuous links between data contained within very large archives, while such archives are composed by structured or non-structured digital documents.

<2> Beyond this obvious yet decisive and practical contribution, Informatics proves instrumental in the epistemological development of the Digital Humanities. New deliverables in the Humanities induce new research methods and concepts and require new interactions between scholars in the Humanities and researchers in Informatics beyond the level of technical assistance and mathematical development. This is specifically true for the case of archives, as both fields are involved, yet by separate means, in a converging process of restoring the original context, circumstances, and/or the experience, of the records considered.

<3> The background of this paper is today's potential (re)creation and (re)shaping of either established, forgotten, unexplored collections of historical documents and records, or of previously non-existing, ex-nihilo and newly invented archives, through computational methods and digital media. The paper considers archives as an object of study that has both greatly evolved and varied throughout the 20th century, in size, form, content, status, purpose, handling or processing altogether, *e.g.*moving from ancient, scholarly or patrimonial valuables to recent, mundane, personal or collective memorabilia; from physical to virtual matter; from archaism to futurism; from paper set to data or even metadata set.

<4> As new types and multimodal shapes of archives emerge it raises the needs to identify what archives are for scholars respectively in the Humanities and in Informatics.

<5> *Archimorphosis* thus intends to be a dialogue between such two researchers trying to bridge the gap between two distinct experiences of archives. While considering recently developed examples of work on audio-visual archival material or other kinds of digital set up within the broad area of Digital Humanities, they may converge in drawing crossovers and parallels, or at least raise shared issues that enlighten the changing status of archives today.

**From archives to archive: reflection on an evolution**

<6> Archives and archive are two distinct acceptations of the word that seem separated not so much by divergent concerns, definitions and even scientific objectives, rather than by academic and scientific methods of approach and practical purposes. Both the Humanities and Informatics resort to the plural "archives" to broadly designate series of documents assembled, variously classified and set aside for different purposes and future uses. Both also use the singular "archive," yet with a different meaning. In the Humanities, especially textual studies, the singular archive has an epistemological and ontological meaning referring to the origin of a work of art in the mind or the personal history of its author. It does not translate into identifiable material documents or features but rather designates the aesthetics and deep underlying specific endeavors in the Arts. From the computer science point of view, an archive refers to digital data stored on a physical medium in order to be preserved during a limited or unlimited period. Such archive can be composed by a simple digital document (text, image, audio, video, or binary raw), or by a huge collection of documents. During the last decade, with the Internet explosion, the development of social networks and the emergence of the cloud infrastructures, digital archive size has exponentially grown, and this tendency is still in progress. A noticeable aspect of the content of these archives is that they may be composed of heterogeneous and unstructured documents as well as homogeneous documents following a very strict format. Origins of such archives are multiple: general public, companies, artists, journalists, scientists who store their digital documents in order to preserve them from mishandling or material failure. These archives can contain documents initially produced under a digital form, but they can also be digitalized forms of material documents. The heterogeneousness and the massive quantity of documents stored in digital archives imply that specific tools must be developed and mastered in order to retrieve relevant information. Also, with new opportunities offered by computers to dematerialize, collect, explore and treat large corpora of information, written, visual and sound documents, that for long could not be grasped or visualized as an identifiable whole, even less processed, we now watch a parallel evolution in the conceptions of archives and archival curation [1].

<7> Collective studies and a current flow of work-in-progress papers describe a shift from a strict notion of conservation and management meant to preserve previously collected materials to a broader dynamic notion: [2] the term "digital curation" is increasingly being used for the actions needed to maintain digital research data and other digital materials over their entire lifecycle and over time for current and future generations of users. Implicit in this definition are the processes of digital archiving and digital preservation, but it also includes all the processes needed for good data creation and management, and the capacity to add value to data to generate new sources of information and knowledge of "creating," generating and assessing new sources of information and knowledge [3].

<8> More specifically, archives in the Humanities were for long seen as mostly unsorted, dusty yet valuable sets of papers, as tangible cultural heritage pertaining to sometimes forlorn pasts; they were essentially riveted on by highly erudite and skilled archivists mastering the techniques of authentication, classification, indexation, and devoted to making physical specimens accessible for potential (yet equally forlorn) future uses. As pointed out by Steven Marcus, "*the Humanities as we think*

*of them, the formal organized study of language and literature, philosophy and history, art and music did not exist in the late 19th century*"; and for a long period of time afterwards," *each(of these subjects was) considered an independent domain, properly organized as its own department of knowledge.*" [4] Thus the use of archives broadly meant resorting to Greek, Latin or Renaissance classics in order to provide a general cultural orientation and guidance, and was almost uniformly oriented toward understanding the past rather than grasping the specific background of works of more recent periods.

<9> Later on, for most of the long 20th century, archives were still seen in the field of Humanities as primary sources of specific and almost encrypted documentation for scholarly inquiries in specialized single disciplines. While cultural studies in the 60ies and 70ies reshaped tendencies within the Humanities, the return of grand theory, cultural anthropology and "*the general notion of the cultural or social construction of knowledge*" [5] meant nevertheless that no trans-contextual validity was seen possible and archives were considered unethical, conservative sets of proofs that could not help establish the genuine sociocultural determination of thoughts and works.

<10> Then came the pioneering cluster of "Humanities informatics" or "Humanities computing," essentially resorting to linguistics in the late 70ies and 80ies; and gradually archives have acquired, especially for the study of written culture at large, an almost reverse status: seen as a series of nondescript collections also open to multipurpose, profane and amateur uses, archives require an interdisciplinary approach mixing the whole realm of methodologies derived from the various branches of the Humanities including social sciences as well.

<11> Challenged by digital research and taken up on the growing interest for a revisited work of art, the restoration of its original context and circumstances of production, archives are not just providing authentic evidences of the past and the means of its understanding; they directly contribute the means of its visible recreation and exhibition (as in virtual museums for example) and the publishing, circulating or displaying of archives thus constitute a valid end-result of research per se.

<12> It is not our purpose to dwell on the recent history of archival work nor of traditional versus digital data collection by scholars in the Humanities, but to reckon that the evolution of the Humanities as a field can be observed through the prism of the emblematic changes in the shape and functions of archival material, which grants the neologism *archimorphosis* a basic epistemic meaning.

<13> *Archimorphosis* underlines the mutation in the medium of communication and display of archives, the consequent reshaping and re-evaluation of its status. Thanks to the potentials offered, the dematerialization of documents is counterbalanced by its visual reconstruction, albeit on screen; so that the issue is not so much the plastic, evolving granularity of the archival objects as its instability, its transitory status leading the document to a permanent metamorphosis and change in format, support, including its re-materialization through printed outputs. Thus a reflection of the metamorphosis of archives rejoins McKenzie's remarks in [6] on the

editorial and bibliographical conflation he observes, the paradox he sees "*in the ease with which new technologies now permit readers to reconstruct and disseminate texts in any form they wish, with few fully effective legal constraints*" and how "*such uncontrolled fluidity returns us to the condition of an oral society.*"

[14] As stated by D.Poulot in,[7] cultural heritage is linked to the representation of values forged by societies, to the ways in which they conceive transmission. In France, the importance of literary value explains why the evolution and construction history of the cultural heritage coincides with the history of texts. Similarly in art history, books set the hierarchies, and the choice of paintings and sculptures in museum collections follows the pattern of indications and quotations by ancient historians in their writings. "*Reality has come to comply with writings.*" [8] It is only since the end of the 80ies [9] in France that the notion of cultural heritage has been reconsidered to include a multiplicity of collections both inert and living, objects of all kinds and all periods of times; the extension of the notion of archives to new media collections requesting new archival methods dates back to the 90ies, as shown by the recent interest for audio-visual archives and their still partial legitimacy.[10]


Figure 1: 1920 children's book illustration

<15> The digitalization of existing archives and the *ex nihilo* development of new digital collections or databases-each being seen as an "archive"-originate new references, imply new concepts and purposes, new systems of organization and display, raising several questions about the status of documents, records or information at large, reconstructed under these new auspices. Among the variations on the theme are the archive as rediscovery, the archive as political act, the archive as argument, the archive as performance[…]. Obviously, digital tools provide a dynamic frame and context to retrieve and display formerly irretrievable objects and artifacts, but even more to exhibit and demonstrate the *relationships* between these objects.

<16> The relationships exhibited can be simple; for example, connect a children's book illustration [11] (*cf.* Figure 1) with the audio version of the popular nursery rhyme [https://biblogotheque.wordpress.com/].

<17> The relationships between archival pieces can also be indirect and less obvious, thus enhancing the impact of digital tools as seen in another example drawn from a different area of knowledge; the bulb in the attached photo, if flatly reproduced in a book, could be seen as a piece of contemporary art. [12] Yet it is a piece of industrial history which even a lengthy written explanation cannot properly reassess.

Figure 2: A bulb: piece of industrial history or piece of contemporary art?

<18> To understand it for what it is, *e.g.* a testimony of past technical progress and industrial know-how, it requires a dynamic refurb and display of the context in which the bulb was manufactured, including added explanations and visual documents about the working space, the social environment, the state and use of electricity at the time and in the geographical area considered, etc. Attempts at a proper recording and transmission of the technical and industrial heritage can sometimes not be completed by traditional means but only digital ones.

<19> A more developed demonstration of relationships can be found in Theaville (http://www.theaville.org) a literary and musical research website (*cf.* Figure 3) and tool dedicated to dramatic parodies of French 18th-century opera [13]. The website has two purposes: one is to reevaluate a vast array of musical works considered in the 18$^{th}$-c. as unofficial; another purpose is to illuminate the hidden and complex mechanisms underlying the composition of these works and the various levels of understanding they require. It does so by connecting music sheets, musical scores, popular versions derived from the parodies, and historic commentaries. The website is constantly edited, and has so far more than 260 parodies with its 2000 corresponding scores, variants, known recorded interpretations, original librettos, sources, links to other resources and archives, and of course explanations, notes, biographies and other works by the composers, etc.

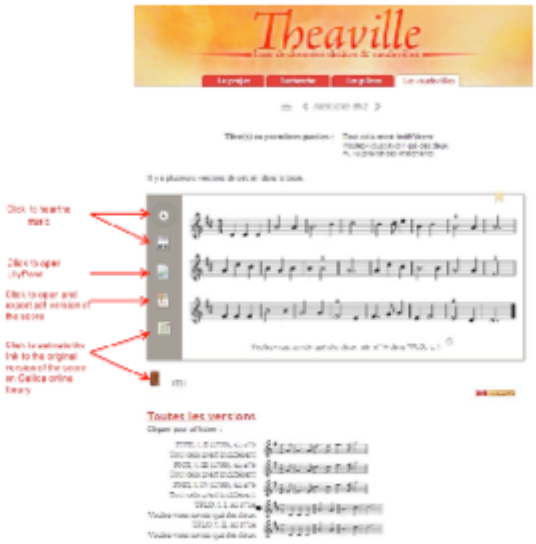


Figure 3: Theaville: a literary and musical research website

<20> Breaking with [biblio-normative] patterns of the past, Web technologies thus provide a flexible, unassuming and user friendly framework of cultural education and information environment at large; they sustain the fluid display of archives or cultural collections at large, which in turn enhances (more or less at will) arrangements by the reader-user; an opportunity and customization of user experience seen as a chain of consequences of user-led, Web 2.0 content creation [14].

<21> The impact of NTIC is therefore especially noticeable on cultural heritage and for recently identified objects, artifacts or documents from the past that were not previously considered and valued as patrimony but became so in the wave of new patrimonial concerns and social sense of responsibilities leading to preserve testimonies of disappearing, or vintage identities, territories, etc. Numerous studies on cultural practices, interactivity and reading within context of saturated digital media and environments [15] also point out that the knowledge resulting from the consultation of digital archives is not so much induced by a scholarly and authorized preconception as it is deducted and reconstructed by the user. While each traditional reader confronts the printed and possibly on screen text with one's specific pace, mode and experience, the printed book still provides a strict rigid frame wherein reading direction and processes (from left to right in Western cultures, from top of the page to bottom, chapter by chapter from the start, etc.) are essentially guided by writing directions. Conversely, digital bodies of documents allow multidirectional approaches and all kinds of purposes, including random browsing, skipping, etc. While the resulting knowledge or content collected is far more unpredictable and hard to evaluate, the open system potentially attracts and empowers, albeit randomly, a wider and non-specialized audience.

<22> Current developments and digital projects in the Humanities based on archives thus face at least a dual challenge. First, a legitimacy issue for archives of contemporary cultural artifacts: what is this fresh set of ephemera with short-term usefulness and popularity? Can it contribute to an understanding of the history of current social trends or living modes? Is this contribution comparable to that of ancient archives, the canonical value of which is guaranteed by time and sustainability? Secondly, and more central to this paper, a formal issue of nomenclature: with the development of multiple digital formats, collections of all sorts need to adjust to the computational or informational definition of archives as this definition undermines or sets aside previous historiographical hierarchies or curation methods and criteria.

**Issues in the shift from material to digital archives**

<23> While *Archives on Fire* could convey a dramatic sense of no return, it is currently argued that the disintegration of traditional representations of archives and cultural heritage is counterbalanced by a correlated and renewed interest for a restoration of theory art, as well as for a reconstitution of the circumstances and means of its conception and production. Moreover, the metamorphosis of archives-the renewal of its form and function in Digital Humanities-correlates with two massive shifts: the movement of all data into digital form; and the creation of new modes of collaboration.

<24> The general reckoning of a change is obviously not a novelty and, since the beginning of the 21st century, scholars in the area of documentation studies have observed how new social and professional practices related to the web and new treatments of information induce a Fredericton notion of data, document, information, knowledge and of their relationships, a deconstruction of working methods, and a mutation in the architecture or structuration of documentation as well as openings to users acting as self-made documentarians. [16]

<25> As stated by D. Cotte, [17] one of the main "change[s] in paradigm" linked to the current generalization of digital documentation relies upon a series of shifts: first, confronting users with techniques and issues in the management of documents that were previously handled by specialists in the field; second, integrating documents or information objects into a broader setting based on computational calculation; thus requiring the users (both scholars and end-users) to rigorously manage, explore and exploit an essentially invisible matter compared to previous handling processes that required a different rationality and sorting out of *visible*documents.

<26> Not only does this shift induce a reconsideration of "*old definitions and distinctions separating publishers, authors, and archivists as well as, occasionally, the separation of producers of scholarly material from its consumers*" [18], but it also increases the argumentative value of the resulting archive: "*Once archived materials are digitized, possibilities for orchestrating access to them and building thematic (and even argument-making) connections with other materials increases exponentially. In structuring the archive, writing its metadata, and designing its*

*interface, the digital archivist is making an argument about the meaning and cultural context of the archive's contents* […]."

<27> As developed by Milad Doueihi in *Qu'est-ce que le numérique?*, *"le contexte change et avec ce changement advient un nouvel ordre, celui des données et de leurs interprétations par les divers modèles de classement algorithmique […] Il ne s'agit point ici de nier la nouveauté de l'accès aux données ni de leur ôter de leur importance et leur pertinence. Mais il reste qu'il faut élucider les manières de lire et d'interpréter ces données et la façon dont ces méthodes pèsent sur nos sociétés. Ce statut emergent de l'information, des données et des nouveaux contextes de leur collecte et de leur divulgation nous invite à revisiter les modèles constitutifs de cette herméneutique de l'information."* [19]

<28> The models that shape the hermeneutics of information are at stake in this paper co-written by a researcher in Informatics and one in the Humanities resorting to different methods of interpretation: the first focuses on the deep structure or architecture of archives; the other is deprived of computational expertise but interested in the potentials of the technology as well as in the openings made to newcomers for a reticulation of archival documents into new patterns or arrangements and the subsequent display of unforeseen knowledge and meanings.

<29> Such a standpoint (illustrating the current intersection of hard and human sciences induced by Digital Humanities) may lead to many questions, including: how to bridge the knowledge gap in computational techniques required to develop consistent digital projects in the Humanities; how to accurately translate research purposes or hypotheses into digital patterns; and conversely how to render computational findings graspable for further observations and research developments in the Humanities. Obviously, these questions will not be explicitly raised here, even less answered, but they implicitly suggest an evolution from the previous ancillary role of Informatics (as a technical yet fundamental support to some areas of research in the Humanities) to a co-understanding and co-construction of digital projects that imply a greater computational literacy for users in the Humanities. This is especially true for the choice and use of metadata, whether structural or descriptive, to support the identification and categorization of data within documents in databases. While inserting metadata and descriptive elements requires a technical expertise, which scholars in the Humanities do not master and find strenuous, defining and listing the metadata requires a global and complementary conception and design both by the informatician, developer or expert in documentation science, and by the scholar in the Humanities who is the likely initiator and end-user of the database.

<30> Digitization of existing archives makes more accessible the information they contain, by the use of relevant algorithms, which allow users from all over the world to remotely request specific queries through the Internet, and allow them to navigate quickly between documents inside an archive. An instance of such digitization of existing archives is illustrated by the Venice Time Machine project. This international project aims to construct a large digital open access database from the archives stored at the State Archives in Venice (Archivio di Stato). Beyond the important efforts to digitize historical Venetian documents by using recent technologies, [20] one purpose of this project consists in

assisting the transcription of digitized texts by analyzing and comparing handwritten sequences of characters in order to detect recurrent patterns. These patterns are then processed manually in order to assign them to words or word sequences. Named entities (proper names or location names for instance) are then detected among these words and a named entity linking [21] is processed, producing a semantic graph of the documents included in the archives. This semantic graph can be perceived as a global semantic structure of the archive, one that is useful to navigate inside the archive.

<31> Pure digital archives (in opposition to archives of digitized documents) are also processed similarly. Even if by nature a digitization process is not required on such data, algorithms applied in order to extract relevant information from these archives are very similar to the ones used to process digitized versions of old material resources. Moreover, the Venice Time Machine project can be compared to the European EUMSSI project, [22] which aims to develop technologies to detect, extract, and aggregate data present in multimedia archives (video, image, audio, text, social content) composed of unstructured information. The EUMSSI project's ambition is to produce tools able to assist journalists in exploring multimedia archives and contents from social networks (tweets for instance). In addition to textual documents, this project focuses on image, audio and video documents that imply the use of pattern recognition algorithms and automatic speech recognition systems. However, all the collected clues are then stored in a database associated with a search platform.

<32> Through these two examples (the Venice Time Machine project and EUMSSI) one can see that traditional archives or more recent ones can now be processed and explored remotely inside large collections of data in real time, allowing the user access to an amount of information impossible to imagine only a few decades ago. Moreover, these algorithms and these semantic representations of the archives may also emphasize links between events, people or locations that were not easily detectable without such numeric tools.

**Information is *on fire***

<33> How can digital archives both convey an accurate transmission of past circumstances and documents while at the same time reconstruct those into new sequences through tools of mapping or indexing meant for new modes of reading and new uses? Is this digital representation a minor or radical alteration of the form and function of archives? Can objects from the cultural heritage be reformatted without a change in purpose and informational value? Are computational methods and codes "light" tools for the adaptation and transfer of documents from the print culture era into a new display of old artifacts? Or does it drastically reshape information for the purpose of a new cultural logic, defined by technological rules and imposing a distinct set of norms and values [23]?

<34> It seems important to better understand how automatic data mining works in order to propose answers to these questions. Data mining consists of the automatic extraction of knowledge from huge amounts of digital documents. Such knowledge can, for instance, be related to named entities

recognition (as defined above), and often focuses on linking different documents stored in the same archive or stored from different ones. This allows us to extract a structure of the data contained in archives, depending on a similarity measure, as presented in Figure 4,in which each link represents a co-occurrence in the same document of person names represented as nodes.

<35> State-of-the-art methods for automatically clustering, linking and classification of textual documents commonly use a vector approach, also known as the vector space model. [24] In this approach, a vector represents a document. For instance, if one assumes that 10,000 different non-function words are sufficient to analyze a set of textual data, each document will be represented with a 10,000-dimensional vector.

<36> Classically, each dimension of this vector corresponds to a word (or a predefined sequence of words), and its numeric value is computed from the frequencies of this word in the targeted document and in other documents of the archive. The classical method to compute such values is called Term Frequency-Inverse Document Frequency (TF-IDF). This calculation is simple, and is related to word frequencies.

<37> Formally, let us consider the number, $N$, of words, $w$, in a given vocabulary, $V$.

<38> In this framework, each dimension of a TF-IDF vector, $\vec{v}$, stands for a word: dimension $i$ is for the $i^{th}$ word in the vocabulary $V$, denoted $w_i$. For each textual document, $d$, in the database, $D$ (the database can be an archive), a TF-IDF vector, $\vec{v}(d)$, can be computed.

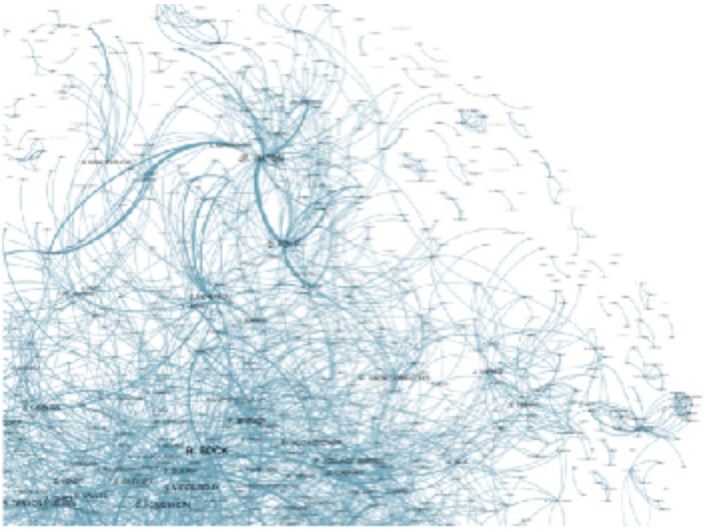

Figure 4: Automatic document linking and visualization

<39> Let $e_i(d)$ be the value of the $i^{th}$ dimension of the vector $\vec{v}(d)$: as explained before, $e_i(d)$ refers to word, $w_i$. The vector $\vec{v}(d)$ will be computed by applying the following formula to its $N$ dimensions $i$:

$$e_i(d) = \frac{tf(i,d)}{idf(i,D)} \quad (1)$$

where *tf*(*i*, *d*) is the Term Frequency of word, $w_i$, in document *d*, *i.e.* the ratio between the number of times word $w_i$ appears in the document, *d*, and the total number of terms in this document.

<40> *idf*(*i*, *D*) is the Inverse Document Frequency, which measures the diffusion of word, *wi*, in the documents of the entire database, *D*. It can be expressed as:

$$idf(i,D) = \log \frac{Total\ number\ of\ documents\ in\ D}{Number\ of\ documents\ in\ D\ with\ word\ i\ in\ them}$$

<41> TF-IDF approach emphasizes words that are specific to a document and relatively rare in other ones. More recent approaches exists to vectorize textual documents, often derived from the TF-IDF approach. The most famous ones are Latent Semantic Analysis or Latent Dirichlet Allocation, particularly effective to emphasize semantic information in the vectorial representation of a document.

<42> Representing textual documents with vector allows us to handle documents through the vector algebra paradigm. So, in order to evaluate similarities between textual documents, similarity measures of vectors are computable. Different similarity measures between two vectors are possible, like Euclidian distance or cosine similarity. Euclidian distance can be expressed as:

$$S_{euclidian}[\vec{v}(d_1), \vec{v}(d_2)] = \sqrt{\Sigma_{i=1}^{N}(e_{i_1} - e_{i_2})^2} \quad (2)$$

while cosine similarity can be expressed as:

$$S_{cosine}[\vec{v}(d_1), \vec{v}(d_2)] = \frac{\vec{v}(d_1).\vec{v}(d_2)}{|\vec{v}(d_1)| \times |\vec{v}(d_2)|} \quad (3)$$

which is the ratio between the scalar product (also known as the *dot product*) of the two vectors and the product of their magnitude.

<43> Both similarities are illustrated in Figure 5, where $d_1$ and $d_2$ refer to two documents. This figure shows also that using a n-dimensional vector representation of a document means that this document is represented as a point in a n-dimensional space. Figure 5 is a trivial case where *n*=2. Actually, clustering a large set of documents in regards of their textual content can be seen as clustering clouds of points; and several mathematical approaches are known to process this problem, [25]like the k-

Nearest Neighbors algorithm [26] or the Hierarchical Cluster Analysis. [27]

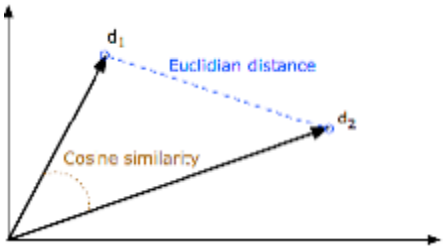


Figure 5: Euclidian distance and cosine similarity between two vectors representing two documents

<44> These examples illustrate how projecting documents into a vector representation is useful to apply automatic analysis based on a mathematical approach. This is also a transformation of symbolic information (words and texts) into numeric information, and similar approaches can be applied to pictures, audio, or video. Since computation power and available data increase exponentially, data mining becomes more and more precise and semantic extraction more and more powerful. [28] These tasks are part of artificial intelligence, [29] which evaluates extremely quickly nowadays. Currently, the deep learning paradigm, [30] which exploits artificial multi-layers of neural networks, allows researchers to make important advances in these applicative fields, with the ambitious objective of creating an artificial intelligence comparable to human intelligence.

<45> Are we now closer or further away from each other than we were at the beginning of this paper? Are we collecting and working on archives with similar purposes and objectives? Probably so, as one seeks to imitate and reproduce human intelligence while the other seeks a comprehensive and evidence-based reconstruction of activities and artifacts. The various processes of Digital Humanities are also an understandable academic response to societal changes, especially in terms of information and knowledge dissemination. And because scholars in the Humanities have the responsibility to look at and produce deeper understandings of past and present, Digital Humanities is a rebranding of a long tradition of scholarship. It is as well a progression of qualitative and quantitative methods through computational methods applied to problems in the Humanities. New archives bear evidence of at least two principles and requests for digitally assisted research in the Humanities: building and making new things and new methods of meanings; and fostering encounters with new and old objects, texts, humans, and non-humans. New archives also engage readers to read outside the traditional boundaries of reading and create the circumstances for new modes of exploratory reading to take place. Thus, scholars in the Humanities seem committed to the powerful tradition of the book: while creating new forms of structured data and new modes for exhibiting archival materials, they still discuss the theoretical and operational changes that may occur "when we think of archives and collections as data aggregations in need of publishing," and state that "The overarching question is: how do new forms constitute

something more than straightforward repositories, publications in wheelwrights?" [31]

<46> Conversely, the example of documents projected in vector representation highlights a reversal, albeit a humanist commitment, to the "counter-mythology which has affirmed the demands of the world, against those of the book." [32] A counter-mythology which McKenzie sees at work in the novel *The Name of the Rose,* where Eco reconstructs "the ingeniously ordered, but labyrinthine, Alexandrian archive, only to deconstruct it again in the old and fearful symbolisms of the library as a furnace. Fire consumes the books." To that apparent, yet bookish counter-mythology, McKenzie opposes Marlowe's Faustus constructing his own versions of selected texts in which he seeks "only a single sense dictating a fixed fate" [33] and yet to his despair experiences variant reading. Pushing the metaphor only slightly further, yet not until it breaks, one could relate Eco's closed-book system library to the traditional model of archival interpretation; and Faustus's encounter with the *"openness to interpretation, the paradox that texts are both closed and open, fixed and flexible, defined by one context only to be redefined in others"* to the digital model of archival interpretation. At a distance from the fact that in both fictions, and in numerous others, books or entire libraries are being burned (and sometimes readers, too), we could say that through digital media, archives may be quasi-literally *on fire* (burned with flames) and consumed. At the same time, the mutual questioning at the background of this paper (as well as the whole Journal's theme) suggests that the figurative meanings of the expression seem to apply both to archives and to information itself, which is *on fire*, that is to say: attractive and doing very well!

**Notes**

[1] The Genesis of data curation and its co-evolution with digital curation is depicted. Carole Palmer et al., "Foundations of data curation: the pedagogy and practice of purposeful work with research data," *Archives Journal* Issue 3 [2013],http://www.archivejournal.net/.

[2] Neil Beagrie, "Digital Centre," *Learned publishing* 17, no.1 (2004): 7-9.

[3] C. Palmer, N. M. Weber, T. Munoz, A.H. Renear, *Archive Remixed, Archive Journal, Issue 3*, Summer 2013: " *As the data accessible in digital environments continue to increase at a rapid pace, sound collection development and description will be essential to presenting data for exploration and discovery. In conjunction with searching capabilities, researchers will greatly benefit from the ability to browse dense and cohesive layers of data sources that not only anticipate researchers' known needs, but also allow them to effectively navigate through and interpret extensive bodies of openly accessible data.*" Ibid.

[4] Steven Marcus, "Humanities from classics to cultural studies: notes toward the history of an idea," *Daedalus* 135, no.2 (2006): 15-21.

[5] Ibid.

[6] Donald Francis McKenzie, *Bibliography and the Sociology of Texts* (Cambridge University Press, 1999).

[7] Dominique Poulot, "L'historiographie du patrimoine," in *L'Histoire culturelle en France et en Espagne*, ed. B.Pellistrandi and J.F. Sirinelli, vol. 106 (Casade Velazquez, 2008), 105-126.

[8] J. Thuillier, Leçon inaugurale. Paris : Collège de France, 13 janvier 1978, 15.

[9] *cf.* Official report by M. Querrien, Pour une nouvelle politique du Patrimoine. *La Documentation française*, Paris, 1982, 5-7.

[10] J. Guyot and T. Rolland, *Les Archives audiovisuelles, histoire, culture, politique* (Paris: Armand Colin, 2011).

[11] Valérie Neveu, "Un nouveau patrimoine écrit en bibliothèque: les fonds de littérature jeunesse," *Les nouveaux patrimoines en Pays de la Loire* (2013): 614-630.

[12] Catherine Cuenca and Yves Thomas, "Constitution d'un patrimoine scientifique et technique contemporain," in *Les nouveaux patrimoines en Pays de la Loire*, ed. Jean-René Morice, Guy Saupin, and Nadine Vivier (Rennes: Presses Universitaires de Rennes, 2013), 669-680.

[13] Published by the CETHEFI (Centre d'étude des théâtres de la Foire et de la Comédie-Italienne), Principal investigator: Professor F. Rubellin, Université de Nantes.

[14] "*Such principles (cf. O'Reilly 2005) included the customization of the user experience, by embracing on-the-fly Web page creation through AJX and other database driven Web technologies which provided an opportunity for users to actively query and select the available information on any given Website. By extension, such technologies also offered an additional opportunity for users to become active as content creators and contributors, which fundamentally altered the vertical alteration between Website providers and Website users. In turn, this potential for user participation and content creation also enabled the emergence of genuine horizontal collaboration between users, and for the formation of user communities[…].*" Axel Bruns,"Produsage," in *Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition* [ACM, 2007], 99-106.

[15] S. Elizabeth Bird, "Are we all producers now? Convergence and media audience practices," *Cultural Studies* 25, nos. 4-5 (2011): 502-516; Anja Bechmann and Stine Lomborg, "Mapping actor roles in social media: Different perspectives on value creation in theories of user participation," *New media & society* 15, no.5 (2013): 765-781.

[16] Ghislaine Chartron and Evelyne Broudoux, "Introduction à la première conférence de Document numérique et Société," in *Document numérique et société* (2006), 7-11; Katell Gueguen, "Traitements et pratiques documentaires: vers un changement de paradigme?" in *Actes de Deuxième conférence Document numérique et société* (Paris, 2008).

[17] Chartronand Broudoux, "Introduction à la première conférence de Document numérique et Société."

[18] Fred Moody, Publishing the Archive, http://www.archivejournal.net, consulted May 12, 2015.

[19] *"The context changes and with the change comes the new order of data and its interpretation by various models of data. Our point here is not to [discredit] its importance and accuracy. But it remains necessary to identify the ways in which such data is being read and interpreted and how these ways affect our societies. The emerging status of information, data and new contexts of collection or dissemination entice us to revisit the models that shape the hermeneutics of information."*

[20] Fauzia Albertin et al., "X-ray spectrometry and imaging for ancient administrative handwritten documents," *X-Ray Spectrometry* 44, no.3 (2015): 93-98.

[21] Ben Hachey et al., "Evaluating entity linking with Wikipedia," *Artificial intelligence* 194 (2013): 130-150.

[22] Jens Grivolla et al., "EUMSSI: a Platform for Multimodal Analysis and Recommendation using UIMA," *COLING 2014* (2014): 101.

[23] The question echoes many similar considerations as rejected. Alice Watterson, "Beyond Digital Dwelling: Practice-Based Solutions for Interpretive Visualization in Archaeology," in *Proceeding of Challenge the past / diversify the future* [Gothenburg, Sweden, 2015], 54: "*In recent years the rising dominance of digital techniques for archaeological three-dimensional surveys and interpretive visualization has resulted in a rapid uptake of emerging technologies without adequate assessment of their impact on the interpretive process and practitioner engagement. Through the observation, exploration and collaboration of various techniques and approaches to visualizing the archaeological record this research moves the current debate forward by challenging common preconceptions and assumptions associated with 'reconstruction'.*"

[24] Gerard Salton, Anita Wong, and Chung-ShuYang, "A vector space model for automatic indexing," *Communications of the ACM* 18, no. 11 (1975): 613-620.

[25] Christopher D Manning, Prabhakar Raghavan, Hinrich Schütze, et al., *Introduction to information retrieval*, vol. 1 (Cambridge University Press Cambridge, 2008).

[26] Yiming Yang and Jan O Pedersen, "A comparative study on feature selection in text categorization," in *ICML*, vol. 97 (1997): 412-420.

[27] Leonard Kaufman and Peter J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*, vol. 344 (John Wiley & Sons, 2009).

[28] Quoc V Le and Tomas Mikolov, "Distributed representations of sentences and documents," *arXiv preprint arXiv: 1405.4053* (2014).

[29] Yoshua Bengio and Yann LeCun, "Scaling learning algorithms towards AI," *Large-scale kernel machines* 34, no. 5 (2007).

[30] Yoshua Bengio, "Learning deep architectures for AI," *Foundations and trends in Machine Learning* 2, no.1 (2009): 1-127.

[31] *Cf. Archive Journal Issue 4*, "Publishing the Archive," Fred Moody (ed.), Spring 2014, http://www.archivejournal.net, consulted May 12, 2015.

[32] D.F. McKenzie, 1999, p. 74.

[33] *Ibid.*, p. 75.

Return to Top»