Writing Documentarity

Arthur Perret
Université Bordeaux Montaigne, arthurperret@me.com

Over the past twenty years, discussions about the nature of documents have often revolved around revisiting the European tradition of documentation. Researchers have taken a new interest in the pioneering theoretical works of authors such as Paul Otlet, Suzanne Briet and Robert Pagès. This has informed our inquiry into the nature of digital documents and data:

> Attempts to define digital documents are likely to remain elusive … Definitions based on form, format and medium appear to be less satisfactory that a functional approach. (Buckland, 1997)

Following this, we have set out to define what digital documents *do* and *how* they do it, more than what they *are* in essence. Borrowing from anthropology, Bernd Frohmann defined documentality as the ability to generate traces (Frohmann, 2012, p. 178). Maurizio Ferraris (2013) also proposed a theory of documentality, which he defined as the recording of social acts in the form of documents. As Claire Scopsi (2018) notes, both approaches relate to the agency of documents. Ronald Day added an important remark: documentality underlines the fact that documents are not simply immovable representations of things but are things themselves, prompting us to action; "documentality is prescriptive, documentation is descriptive" (Day, 2018, p. 8). It should be noted that this discourse on the use and the agency of documents draws directly from both Otlet and Pagès:

> Material things themselves (objects) can be considered as documents when they are taken as discernible elements, directly from studies, or as evidence in a demonstration. This is "objective documentation" or "automatic documentation." (Otlet, 1934, p. 217)[1]

> An anonymous Egyptian mummy, a gorilla in a cage, a piece of Spath … in this case the document transmits information about itself. It is an "auto-document." (Pagès, 1948, para. 46)

Documentality is not to be confused with documentarity. The words are almost identical and as concepts they come from the same functional approach to documents. However, they take a different path. In his recent book on documentarity, Day (2019) frames it as a philosophy of evidence built upon the history of inscription. Here we offer additional insight into both elements of Day's proposal—evidence and inscription—by discussing previously unaddressed but relevant works from the French and American scientific literature. This opens new avenues for both theory and experimentation.

Documentarity as a quantifiable quality

Ronald Day's book *Documentarity* is the product of interdisciplinary theoretical work, at the intersection between ontology and documentation. The central

---

[1] All translations from French works are the author's own, except when mentioned otherwise.

concept is defined as a philosophy of evidence based on inscriptional technologies of judgment[2].

The basis of this work is philosophical. Day leads with a close reading of Martin Heidegger's critique of technoscience. The frame of Day's proposal is poetic in the sense of Heidegger: it explores expression not as anthropocentric engineering but as an interaction between affordances. He further develops his point by borrowing from Bruno Latour's pragmatic approach to substance and inscription. This helps him formulate a view of information-as-process, a *poiesis* of which an entity is the focal point. Finally, Day draws from Rom Harré's distinction between dispositions and affordances to explain the balance between internal and external powers of expression.

From this, Day derives a practical framework. He proposes a distinction between two forms of documentarity: a strong documentarity, rooted in *a priori* categories and ideal reference; and a weak documentarity, produced *a posteriori* by empirical sense. The tension between the two is somewhat resolved in the case of computer-based information technology, which Day closes the book on. These last pages differ from the rest: instead of delving deep into a comparison between 2 or 3 examples, Day reviews more briefly a wider array of phenomena to which he applies the strong-reference/weak-sense approach. His remarks are insightful but they do not quite bring about the shape of the digital *poiesis*, the form of information-as-process in the computer paradigm.

There are two significant occurrences of documentarity in literature prior to Day's book, which provide us with an opportunity to address this. Before it was used in relation to documentation, the word documentarity first came up in film studies, specifically on the topic of documentary films. It was defined as the answer to the following question: "*qu'est-ce qui fait document ?*" (Gaudreault & Marion, 1994, p. 13). The translation of this sentence is tricky, because the French verb "*faire*" is used in a secondary sense which is closer to "seem" than "make": "*donner une qualité, un caractère, un état à.*" (to give something a quality / character / state of)[3]. Consequently, we should not translate Gaudreault and Marion's question literally ("what makes a document?"). Instead, a better, more accurate (if not elegant) translation could be: what is it that makes something seem documentary?

> An image always presents a greater or lesser degree of resemblance with the object which it is modeled on, and thus can always claim to « seem documentary »[4]. This claim to a greater or lesser « documentarity » is dependent on the medium …

---

[2] The *Oxford English Dictionary* defines "evident" as: "Clear to the understanding or the judgement".

[3] From https://cnrtl.fr/definition/faire, II. C.

[4] In French: *faire document*.

> Photography has, ontologically, a high degree of documentarity … The degree of documentarity of a medium depends on its ability to show a greater or lesser number of indices of reality. (Gaudreault & Marion, 1994, pp. 17–19)

According to this, documentarity is at the same time a quality or property—in the spirit of the polysemous German word *Eigenschaft*—and a quantifiable thing. This is also the case in the second occurrence of the word, which can be found in the works of Stéphane Crozat. He defines documentarity as "a measure of what a content enables through a writing contract based on its documentary properties" (Crozat, 2016). His definition is completely unrelated to the previous one and uses an entirely different theoretical framework—redocumentarisation (Pédauque, 2007). However, it expresses roughly the same idea: documentarity is a property on the basis of which we judge information. By putting the word "measure" in the front of his definition, he echoes indirectly Gaudreault and Marion's "degree of documentarity", suggesting that it is a quantifiable quality. In both instances, the concept of documentarity translates the fact that media are involved in processes of communication; it fits within a theory according to which documents are information recorded to be transmitted, and in which the question of their value is largely tied to their eventual interpretation. Compared to documentality and to Ron Day's documentarity, the focus here shifts from expression to reception.

The role of writing in document theory

How do we assess documentarity? As Otlet noted, "the smallest document is an inscription" (Otlet, 1934, p. 43). This is a simple but powerful statement which directs us to inscriptional technologies. This course of inquiry is not new: in his review of the links between semiotics and information science, Julian Warner concluded that "documents and computers are unified, and differentiated, by the presence of writing" (Warner, 1990, p. 28), calling for a deeper exploration of this idea. Day himself introduces his book with the observation that "too little attention has been paid to the aesthetics of information" (2019, p. 3). In his study of the relationship between language, speech and writing, Jack Goody demonstrated how lists, tables and recipes enable us to do more with our brain—what he called writing as a technology of the intellect (Goody, 1977). Applying his concept to networked computing, others have discussed what it could mean in a broad perspective, however without actually delving into the fabric of writing itself. To examine the way documents and data become manifest in digital form, we need to look at how signs and media have evolved too.

The theory of "screen writings" (*écrits d'écran*) (Jeanneret, 2005), which applies the semiotic approach to computer-based communication, aims to research modern textuality. It is notable for its study of writing programs through the concept of architext, which is loosely defined as a category of tools which allow us to write on computers. The wordplay between architext and architect is

intentional: it leads to a critique of the way software can be designed to control expression.

Because it was used mostly in the context of Graphical User Interfaces (GUI), there is room for the concept of architext to grow and to inform the issue of documentarity. If we look at widespread file formats designed to carry text, we find they often use a hierarchical tag system expressed in one or another markup language (ML) e.g. Web pages are written in HTML (HyperText ML) and Word files in a format based on XML (eXtensible ML). By definition, GUI do not display markup; the "document" we see is not what is stored in the file system but the product of rendering. Samuel Goyet (2017) applied this logic to Application Programming Interfaces (API), a critical mechanism for building Web pages. By shifting the focus from display to code, he exemplified how documents are built dynamically from reticular writing, organized and structured through markup and links. In his view and others' (Collomb, 2017), this creates the opportunity to open the definition of architext to code. But to do this, we need to move beyond what Clarisse Herrenschmidt describes as the "simulacrum" (rendering) of GUI and closer to what she calls "simulation"—visible, algorithmic inscription (Herrenschmidt, 2007, p. 398).

How documentarity is written

If we read marked up text in a plain text environment, we can distinguish two categories of signs. In the first category are signs for which there is no equivalent in the world of pen and paper e.g. temporary markers of interaction such as cursors and selection highlighting. In the second category, we recognize alphanumeric characters and punctuation marks, but the latter call for deeper examination. Typography expert Roger Laufer considered that writing and printing brought authentic, significant semiotic inventions—enough to warrant new terminology. He coined the term *scripturation* to properly address this and distinguish "marks of enunciation" from signs that match the inflexions of spoken language. An exclamation mark belongs to punctuation but dashes and brackets belong to scripturation. By inventing this word, Laufer wanted to draw our focus to the role of these inventions, especially the way they signal various levels of structure:

> This is the generic term I propose to designate all marks of enunciation, handwritten and typographical … Non-punctuation *scripturation* is intra- and supraphrastic: it refers to the most general divisions of documents, such as parts or chapters, in the table, paragraph, bracket, hyphen, bracket, italics. (Laufer, 1986, p. 75)

Scripturation is enunciation made evident; it is the practical and intellectual basis of markup. In fact, the Generalized Markup Language (GML) invented at IBM in the 1960s was a port of editorial codes (e.g. "Body" for "Times 12pt justified") onto computers in the form of tags and delimiters. These made extensive use of

scripturation and punctuation marks—from brackets, dashes and backslashes to colons, carets and apostrophes—and this legacy is present in all markup today. It can be seen in languages designed to carry data in general (XML, JSON) or text in particular (HTML, Markdown), in typesetting languages (LaTeX), in stylesheet languages (CSS, CSL), etc. The fact that the same set of signs is used to store, transport, structure, style and display information shows us that there is indeed a unifying logic to computing, writing, documents and data. Delimiters were in use long before the computer, the printing press or the alphabet; therefore the encoding of data and documents is tied to the same long history. Markup belongs to technologies of the intellect in the same way that lists, tables and graphs do. This leads us to propose an alternative definition of the architext as a technology of the intellect which organizes enunciation; it is scripted text—*une écriture de l'écriture*.

It could be said, syllogistically, that since architext is the way we organize enunciation and that documentarity is a property of documentation, documentarity is enabled by way of architext. However, documentarity is not any characteristic of documentation: it defines the very fact that we call documentation that way. So whenever architext can be applied as a framework to explain the enunciation of something we call document or data, it overlaps with documentarity. This overlap makes it easier to understand what may affect this quantifiable quality. Indeed, any process of documentarisation or editorialisation has to do with the architext: humans and machines can read and write architext, and use it to create, combine and disseminate information. Digital products of document acts and knowledge organization are architextual. We can simply read it to assess the structure, the presence of data and metadata, the formatting rules that apply to it, the links to other documents, etc. That is, if architext is readable. Unfortunately, the technological mediations of read/write processes are not always as simple.

The texture of enunciation

The concept of architext was originally tied to the study of computer writing in the context of software development. However, its authors quickly moved on to rich text and media editing. The technological mediations are very different in these two contexts and explain in part why they did not associate architext and code, something that has only been done very recently (Collomb, 2017; Goyet, 2017). Interestingly, the history of the word architext itself provides us with insight here, through a short historical detour.

"Architext" was borrowed by Yves Jeanneret and Emmanuël Souchier (1999) from French linguist Gérard Genette. The meaning was changed in the process and most people who quote their use of the word are unaware of this broken filiation. At the end of the 1970s and the beginning of the 1980s, Genette had an interesting exchange of sorts (by interposed publications and footnotes) with his

American counterpart, Mary-Ann Caws, over their respective use of similar terms in very different meanings—Genette used architext while Caws used architexture.

> Architexture is meant, in brief, to stand for the building of the text as it is seen and is formed with the reader's collaboration, special attention being given to the surface of the building material, its *texturality*. (Caws, 1981, p. 10)

This definition was written in the context of poetry: according to Caws, the length of the line, the rhyming and stylistic effects (such as metaphors) all arrest the eye when we read. They form as many bumps and ridges on the surface of the text while it takes shape during our interaction with it. Now, coming back to the architext in the sense we give in the context of this paper (scripted text): if scripturation is the texture of enunciation, it becomes crucial that we be able to *sense* it. As Herrenschmidt wrote, "there is writing when, the writer being absent, another person can read and know the contents of the text" (2007, p. 75). Markup can be opaque and/or obfuscated. This raises a question: can we always properly assess documentarity?

Any interface to a database is a good example to comment on the various ways documentarity can be more or less well sensed, let alone measured. As an example, we will briefly discuss the following screen capture (Fig. 1). It shows 4 different ways one particular dataset can be interacted with. The test was conducted on Isidore (https://isidore.science/), a search engine which harvests records from other databases in French humanities and social sciences and enriches their metadata.
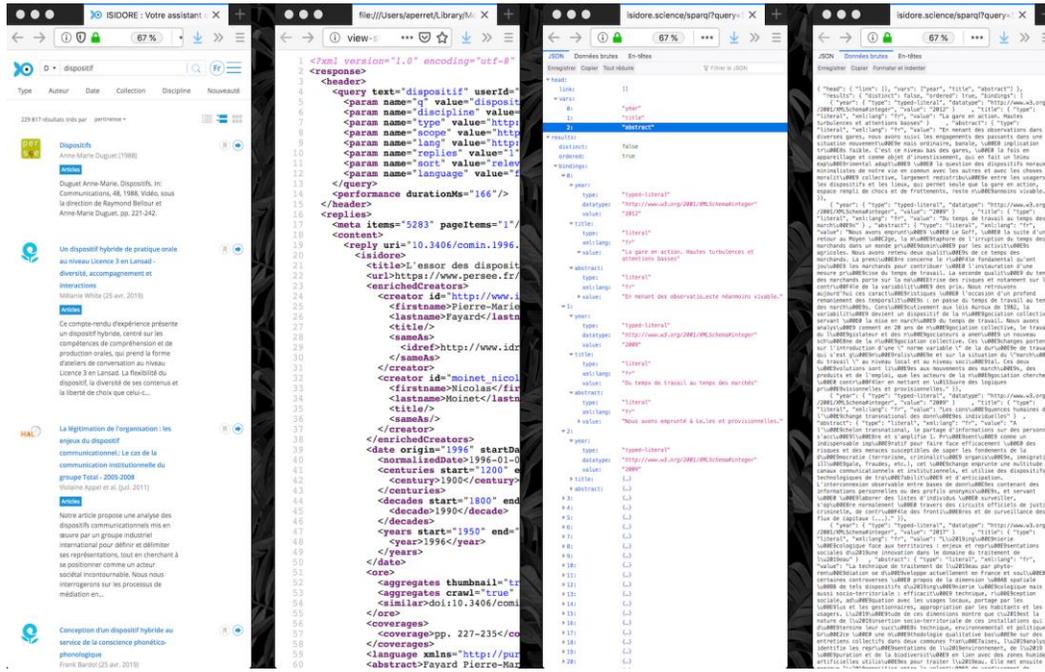
Fig.1 - From left to right: web interface; XML; JSON with interface; raw JSON.

A query ("dispositif") is made. The 1st window shows the results directly on the website. The interface is entirely organized through web technologies: structured content (HTML), stylesheets (CSS) and automation (Javascript). They are set up, coordinated to determine what we see: the amount of information, its hierarchy, its look and feel. This is editorialisation: here we are the closest to the printed page paradigm, where everything we see has been thought through. By contrast, the 3 other windows show the same results but in their entirety, as output from the API (for XML) or the SPARQL endpoint (for JSON). The hierarchy is the same, but the amount of information is drastically different, and its appearance even more so; in these cases, we are closer to plain documentarisation. The browser used to display these results is the same in each case (Mozilla Firefox). It does not handle all types of markup in the same way: XML is colored based on its syntax; JSON is presented by default through a unique interface (which has nothing to do with the website but is a feature of the browser); raw JSON is "minified", which means it is stripped of spaces and indentation.

This all affects the perception we have of the information. Raw results are difficult to navigate; but the web interface shows us very little by default. The browser offers a useful interface for JSON data; but the website has a friendlier design. Whatever choice we make, documentarity will be increased or diminished. A simple example such as this one shows us that documentary quality varies based on documentarisation, editorialisation and reception—all dependent

on the underlying technological inscription that is the architext and on the way we receive it. Such exploration suggests that the line between the theories of writing and the theories of documentation is very thin.

Conclusion

Documentarity brings something different to document theory. While it does touch on the essence of what a document is, it does not require us to ascertain whether something is *essentially* a document; instead, we may simply assess the degree of its documentarity. In the same way that Otlet spoke of "substitutes of the book" (Otlet, 1934, p. 217), we might speak of "substitutes of documents": digital objects which challenge our current conceptions of the document but fit within documentation as a science and a field of practices.

The distinction between strong-reference and weak-sense documentarity introduced by Day is a powerful tool to explain the logic behind information technologies. However as a framework to understand the digital paradigm, it needs a few more beams. The reason why Day does not need to elaborate on the materiality of documents when discussing documentarity in the context of Otlet and Briet is that we know it quite well from decades of scholarly work; this is not the case for digital materials. There is no equivalent yet in breadth or depth of the work done for example in media archaeology. In France, the field of mediology produced interesting preliminary works but is somewhat dormant (cf. Debray, 2000 and the *Medium* journal). More recently, techno-semiotics have been favored by a new generation of researchers in information science, with promising results. Our description of the architext as a tool to characterize the shape of enunciation participates to this effort.

The architext helps us understand the dispositions and affordances of digital documentarity by showing that information-as-process is no more an abstraction in this context than it is for analog media: it is supported by technologies of inscription which we need to describe (scripturation, markup) because they inform our view of information experience. The importance of aesthetics as evidenced by Day suggest that more interdisciplinary work on this topic has yet to come.

Frohmann suggested that information science should draw from a more diverse range of disciplines and experiment with new concepts:

> The temptations of a Theory of Everything are often irresistible. But there are other approaches to documentation . . . forging concepts in a Deleuzian spirit, with more concern for what they do than for what they mean or represent (Frohmann, 2009).

The usefulness of such experimentation lies in the way it shifts our perception of things, introduces news ideas, dislodges pre-conceptions. It fits within a science which acknowledges that it is a permanent work-in-progress: not a Theory of Everything but intellectual tools to be tested and debated.

References

Buckland, M. K. (1997). What is a « document »? *Journal of the American Society for Information Science*, *48*(9), 804-809. DOI : 10.1002/(SICI)1097-4571(199709)48:9<804::AID-ASI5>3.0.CO;2-V

Caws, M. A. (1981). *The eye in the text*. Princeton : Princeton University Press. http://archive.org/details/eyeintextessayso0000caws

Collomb, C. (2017). Faire compter les machines. *Communication*, *34*(2). DOI : 10.4000/communication.7327

Crozat, S. (2016). Proposition : principe de documentarité. *As we may...* http://aswemay.fr/co/010013.html

Day, R. (2018). Auto-Documentality as Rights and Powers. *Proceedings from the Document Academy*, *5*(2). https://ideaexchange.uakron.edu/docam/vol5/iss2/3

Day, R. (2019). *Documentarity: Evidence, Ontology, and Inscription*. Cambridge : MIT Press.

Debray, R. (2000). *Introduction à la médiologie*. Paris : Presses universitaires de France.

Ferraris, M. (2013). *Documentality: Why it is necessary to leave traces*. Traduction par R. Davies, New York : Fordham University Press.

Frohmann, B. (2009). Revisiting « what is a document? ». *Journal of Documentation*, *65*(2), 291-303. DOI : 10.1108/00220410910937624

Frohmann, B. (2012). The documentality of Mme Briet's antelope. Dans J.

Packer & S. B. C. Wiley (Éds), *Communication Matters: Materialist Approaches to Media, Mobility and Networks* (pp. 173-182). London : Routledge.

Gaudreault, A., & Marion, P. (1994). Dieu est l'auteur des documentaires…. *Cinémas : Revue d'études cinématographiques / Cinémas : Journal of Film Studies*, *4*(2), 11-26. DOI : 10.7202/1001020ar

Goody, J. (1977). *The Domestication of the Savage Mind*. (S.l.) : Cambridge University Press.

Goyet, S. (2017). *De briques et de blocs. La fonction éditoriale des interfaces de programmation (API) web : entre science combinatoire et industrie du texte*. Thèse de doctorat. Université Paris-Sorbonne, Paris. https://tel.archives-ouvertes.fr/tel-01665406/

Herrenschmidt, C. (2007). *Les trois écritures : langue, nombre, code*. Paris : Gallimard.

Jeanneret, Y. (2005). Sémiotique de l'écriture. *Canal-U*. https://www.canal-u.tv/video/fmsh/semiotique_de_l_ecriture.28783

Jeanneret, Y., & Souchier, E. (1999). Pour une poétique de l'écrit d'écran. *Xoana*, (6/7).

Laufer, R. (1986). L'énonciation typographique. *Communication et langages*, (68). DOI : 10.3406/colan.1986.1762

Otlet, P. (1934). *Traité de documentation. Le livre sur le livre* (3ᵉ éd.). Bruxelles : Les Impressions nouvelles (2015).

Pagès, R. (1948). Transformations documentaires et milieu culturel (Essai de documentologie). *Revue de Documentation*, *15*(3), 53-64.

Pédauque, R. T. (2007). *La redocumentarisation du monde*. Toulouse : Cépaduès-Éd.

Scopsi, C. (2018). The Documentality of Memory in the Post-Truth Era. *Proceedings from the Document Academy*, *5*(2). https://ideaexchange.uakron.edu/docam/vol5/iss2/4

Warner, J. (1990). Semiotics, information science, documents and computers. *Journal of Documentation*, *46*(1), 16-32. DOI : 10.1108/eb026850