

WHY THE DOCUMENT IS IMPORTANT . . .
AND HOW IT IS BECOMING TRANSFORMED

From the Latin *documentum*, the French word *document* appears occasionally in the thirteenth century, most often in the plural form *documens* or *documenz*. The *Dictionnaire du Moyen Français*¹ gives two meanings:

- A. Lesson, teaching, . . .
- B. Writing that serves as evidence.

Thus the term has since the Middle Ages come to acquire two complementary meanings that have established themselves over the centuries along the dimensions of transmitting and proving. In the Middle Ages the *documens* are first of all lessons that are not necessarily put down in writing—or *documented*. The Latin *doceo* means ‘teach’; *documentum* is the act of teaching.

The second meaning, of proving, or evidence, refers to titles, often title deeds. The English *record*, literally a ‘recording’, is often translated into French by *document*, which accounts well for this function. Here is what the Oxford dictionary says of its etymology:

Middle English: from Old French record ‘remembrance’, from recorder ‘bring to remembrance’, from Latin recordari ‘remember’, based on cor, cord- ‘heart’. The noun was earliest used in law to denote the fact of being written down as evidence. The verb originally meant ‘narrate orally or in writing’, also ‘repeat so as to commit to memory’.

We still have in English the expression: ‘learn by heart’ similar to the French ‘apprendre par coeur’.

Thus, according to their etymology, the French *document* and the English *record* have almost the same meaning, though they come from two different Latin words: the first emphasizes transmission and the lesson, the second memory and evidence.

“Why the Document Is Important . . .
and How it Is Becoming Transformed” by Jean-Michel Salaün,
The Monist, vol. 97, no. 2, pp. 187–199. Copyright © 2014, THE MONIST, Peru, Illinois 61354.

The modern concept of a document combines the two senses, probably as a result of the slow crossbreeding of (confidential) legal practices and nobler (and more open) monastic transmission practices. It also results, perhaps even more surely, from the transformation of the relationship to truth which is no longer revealed, preserved, and handed down by King and Church, but constructed through reason and demonstration.

In any event, the term ‘document’ does not seem to have been much used until the eighteenth century, though already in this century the public sphere is marked by an impressive epistemic infrastructure of museums, academic libraries, and scholarly journals, which serve the presentation, classification and comparison of documents arising from scientific activity. The word ‘document’ actually becomes prevalent only in the nineteenth and especially the twentieth century, when the meaning is established of a privileged vehicle in the circulation of knowledge, in the dissemination of technology, and also in ensuring the stability of organizations and the development of trade. Therefore, it is likely that its contemporary success stems directly from the scientific and industrial revolution, which needed powerful tools to transmit, convince, and prove.

The growing popularity of the word reflects this vast movement. But, in France, it is not until the ninth edition of the *Dictionnaire de l'Académie*—dated 1992 (!)—that its definition is expanded beyond two lines. Yet today we are overwhelmed by documents, in our pockets and on our furniture, in our workplaces and in our home, in public spaces and even more all over the web. The whole social sphere is affected: family, school, government, the organization of work, industry and commerce, recreation, culture, the definition of states, the development of public space, international exchanges and, of course, technology, science, and law. Documents are involved in the regulation of all the workings of society, often today including our privacy. We speak of ‘documents’ about everything, to the degree that the word is somewhat emptied of its substance.

1. First Definitions

The first serious reflection on the term ‘document’ is probably that of the Belgian Paul Otlet (1868–1944):

The Book [or document] thus understood has two aspects: a) it is primarily a work of man, the result of his intellectual work; b) but multiplied in many

copies, it also presents itself as one of several objects created by, and able to affect, civilization.²

This definition emphasizes a double multiplication of documents, into a multiplicity of types, and into numerous copies. It relates to published documents.

During the 1950s in France, Suzanne Briet, in a small passionate book, broadened and deepened our reflections on this topic by proposing the following definition:

[A document is] any concrete or symbolic sign, stored or recorded, in order to represent, reconstruct or prove a phenomenon either physical or intellectual.³

This definition is very broad: any object carrying information can be a document provided it is included in a documentation system. But where Otlet emphasizes transmission, Briet emphasizes the factor of evidence in the broad sense of faithful representation. To document is to reflect the reality of some thing, event, phenomenon, or process of reasoning. The two original functions merge increasingly in modern scientific publications, where the peer-reviewed journal article is used both to transmit and to prove.

By most contemporary standards the document is an object (physical or electronic) on which information is recorded. It would thus have two dimensions, the medium and the content. But this dual presentation is insufficient: it obscures the social function that lends the documentary function to both medium and contents.

A good illustration of this ambiguity can be found in the legal framework for information technology of Quebec.⁴ Quebec law is interesting in this respect because, it tries to define a document beyond the medium it uses by paying attention to information. We can read in Article 3 of the 2001 law this definition:

Information inscribed on a medium constitutes a document. The information is delimited and structured, according to the medium used, by tangible or logical features and is intelligible in the form of words, sounds or images.

On the face of it, this passage defines a document only in terms of its medium and of its contents. These contents, moreover, are viewed as independent of the medium. But the appearance is deceptive. On the one hand, it is precisely because the document has a function—that of transmission of evidence—that we need a law to define it. We must, indeed, be sure that the object we are talking about will perform this function in the new

digital environment. On the other hand, it is indeed because the content can pass from one medium to another that Quebec has tried to define in law the link between one and the other to ensure that the documentary function is preserved.

In France, a network of researchers writing under the pseudonym Roger T. Pédaque⁵ has proposed in contrast a three-dimensional grid—form, content or text, and mediation—summarized mnemonically by the three past participles: Seen, Read, Known.⁶

2. *Seen: the Form*

The first dimension, that of the form, which means: of what is seen, is anthropological. This is the relationship of our body and our senses to the document as object, regardless of the meaning it conveys. A document must be immediately comprehensible as such. An informed reader will be able to identify the publisher of a book, or the series to which it belongs, even from afar in a library. Its cover, its title, its author will allow him to classify it as belonging to a particular genre. The logical structure (of pages, chapters, table of contents) is a familiar one which allows him to browse the book without getting lost.

This skill in detecting and consuming documents is essential to our daily lives and is of course not exclusively related to the form we call *book*. We can easily recognize a passport among dozens of other objects on a table. It is also important that we can easily identify invoices, contracts, correspondence, and so forth. Conversely, if we are not familiar with the look of a certain document—if it does not fit into an environment that we know well enough that we do not get lost—then it will be difficult for us to detect, and may simply not exist for us. For others, in contrast, it will have an immediate significance.

Pédaque summed up this first dimension by means of the equation:

$$\text{document} = \text{medium} + \text{inscription.}$$

For printed documents, inscription and medium are consolidated. The solution of the equation in this case therefore comes down to a simple matter of perception. Matters are more complicated in the case of an audiovisual document, for here both recording and playback require special equipment. The document, understood along the dimension of what is perceived,

needs to be reconstructed in a form isomorphic to the earlier recorded signal. Decoding is a transposition from one form to another.

For digital documents, the equation is more complicated still. Here several different documents can be returned from the same digitally stored items differing significantly in form. This is why the above-mentioned legislation in Quebec insists on the permanence of the structure of information in a document.

In its first dimension—the dimension of form—the document is only of the nature of a promise. Even without knowing its content, we already know that it is a document and we sense that it will be useful for us, whether or not it will be interesting or entertaining. But a promise must be kept, otherwise it avails nothing. Seeing a document is not enough; we should also be able to view its content, even if this potential is not always actualized.

3. Read: the Text

The second dimension, that of contents or text, is an intellectual dimension.⁷ What is at issue here is the relationship between our brain, with its reasoning skills, and the contents of the document, regardless of how this content is represented, be it fixed in writing or a matter of moving pictures or sound or a combination of all of these. Through our senses, our brain decodes an inscription in order to interpret the representation it offers. In this dimension signification is in the forefront. Decoding the inscription can be achieved by us, by a machine, or by both. In the case of the most traditional document, one carrying a written text, decoding is done by the reader. Such decoding requires that the reader, usually in his childhood, has been submitted to a long apprenticeship in order to be able to read. We can often identify a document written in a language or a script unknown to us by the first dimension, its form; but it is impossible for us to decipher it, to know its contents, except perhaps imperfectly and by analogy.

It is to account for this dimension of the text that researchers have proposed the notion of a reading contract, and they have shown in particular that in literature this notion underlies the idea of classification by genre.⁸ For a text to belong to a genre it must follow certain rules of writing. It is the familiarity of these rules on the part of the reader that allows him to enter the world of the author.

Computer technology has taken us one further step by allowing software code to perform calculations and logical manipulations. Digital docu-

ments can not only be decoded by machines, but “interpreted” by them. This opens the door to many new applications in automatic language processing and knowledge engineering, including attempts at automatic translation, as well as search engines and Semantic Web applications.

Today we scroll through documents, we navigate, we consult, we search on the web. We copy and paste rather than focusing our attention on long texts. The text is hereby at least partially deconstructed and new reading contracts are gradually establishing themselves and new genres are thereby being defined: blogs, YouTube videos, tweets, and mixtures of audio and video contents (mash-ups). Beyond this, algorithms can now reconstruct documents on the fly. We have not as yet grasped the consequences of these new kinds of reading contracts between readers and machines, or of what Alain Giffard has called the new ‘industrial reading.’⁹

4. Known: the Mediation

The third dimension is that of mediation. Whatever its form and content, the document must have some social function. We have seen that the two original social functions of transmitting and proving have evolved, both through merger and through expansion. Today we say ‘to inform,’ but the sense of the verb is not precise. Nonetheless, our initial distinction is helpful because it reminds us that the two terms do not refer to the same universe. To transmit is conjugated in space and time, while to prove is conjugated in truth and trust.

When viewed along the dimension of transmission, the purpose of a document is to transmit a content to others, beyond the here and now, using some form or format. The document is associated with some distribution system through which it is able to serve as a medium which allows some formatted text to escape the narrow bounds of space and time and of the private circle of those with whom we communicate directly by means of speech. The document is a memory object: some information (text) is recorded on an object (form) in order for it to be transmitted or referred to and thus to become known in some other place and at some other time. To paraphrase Stanislas Dehaene and Michel Serres,¹⁰ the entire documentation system thereby forms our external memory.

Along the proof dimension the system works only if the document is able to call in aid a contractual framework in which all the links in the chain are organized properly and all interests preserved. The term “contract” here is used in a less metaphorical or implicit sense than in the reading

contracts we can identify along the form and content dimensions. “Contract,” now, can have legal or financial implications, and often involves obligations, as when an invoice leads to payment.

The relationship to the past differentiates a document from a performance—since the latter occurs at a time and in a place where it is listened to or watched by an audience. This differentiates a document also from a conversation where the interlocutors are present. For performances and conversations, live television and radio shows and the telephone helped break the barrier of space. However, to break the barrier of time it was necessary to use recording devices and thus to turn performances and conversations into documents. We see once again that digital technology, and now especially the web, have significantly changed our relationship to time and space. The distinction between performance and document fades, as does that between communication and information and between conversation and publication. We record, discuss and publish with one click.

5. Integration

Each of the aforementioned dimensions of form, text, and mediation has its own logic, which cannot be reduced to the logic of the others. And yet none is completely independent of the other two. A document must integrate and coordinate the three dimensions. Its anthropological, intellectual, and social modalities must each perform their role when taken separately; but they must also be consistent with one another.

Pédauque, who first suggested this partition into three dimensions, summarizes his proposal by pointing out that:

in each case we have stressed the idea of a reading contract, expressed by readability in the first dimension, by understanding in the second and sociability in the third. It seems likely that this three-faceted contract expresses the reality of the notion of document in all its aspects. A document is in the end a contract between human beings whose anthropological (readability—perception), intellectual (understanding—assimilation) and social (sociability—integration) skills are, at least partly, the basis of their humanness and of their ability to live together. From this perspective, digital technology is merely a modality of multiplication and evolution of these three kinds of contracts.¹¹

6. A Proposal

Based on these findings a provisional redefinition of ‘document’ might read as follows:

A document is a trace used to interpret a past event through a reading contract.

We find here all three dimensions:

material: with the reference to a trace (that is seen)

intellectual: with the reference to an interpretation (that is read),

transmission: with the past event (known), and the necessary social construction involving some contract.

The notion of trace broadens the definition of ‘document’ to include many different kinds of items, exactly along the lines proposed by Suzanne Briet. To use her examples, both a star in the sky and an antelope can be documents, provided only that they are knowledge-items embedded in an information storage and retrieval system providing them with what we now call ‘metadata’.

Anything can be a document in this extended sense; but not anything *is* a document. The thing in question needs to be singled out, identified, and embedded in some documentation system: it has to be converted into a document in such a way that it can be interpreted in the way appropriate to the system to which the trace belongs, whether legal, literary, scientific, customary, commercial, and so on.

Finally, talking of ‘reference to a past event’ does not necessarily imply that this relationship is made explicit in the document. It implies merely that the document establishes some link between today and something that has happened in the past of which it is a trace. The trace can be a deliberate reference that is built into the text; or it can be a mere hint that may or may not be identified as such by the reader. A document is a way to regain our past and to rebuild it in accordance with our present in order to guide our future. Even a work of fiction reflects its time—the time of authorship as interpreted by the author—even if its content is imaginary and is happening at another time in history. A fiction is a story handed down by its author to posterity. Its text is a trace of an act of writing, and thus of a past event (the writing occurred in a given temporal context). The text can thereafter be transmitted to readers who will interpret it on the basis of their own context at another time in the future.

But the above definition is still flawed, in that it does not account for the essential feature of ordinary documents that they are easily manipulable according to familiar protocols. It is this feature that has made possible their inclusion into document systems and thereby ensured their success. This feature explains the reproducibility of documents, their plasticity, and the multiple familiar ways in which they are treated from day to day. If, as Briet argued, many different sorts of object can become documents, then in many cases the documents in question—footprints, or cave paintings, for example—would be *sui generis* and not easily manipulable according to recognized protocols. Our definition still needs to do justice to the fact that the document in the most ordinary case is a text, a representation in the form of a repeatable prototype in an easily manipulable medium such as paper or, now, a digital file.

The definition proposed by the famous Indian librarian Ranganathan emphasizes precisely this feature of easy handling:

A document is an ‘embodied micro thought’ on paper or other material, fit for easy physical handling, transport across space, and preservation through time.¹²

Ranganathan’s definition, however, has the opposite flaw, that it does not take account of the fact that objects of any kind are potential documents. In addition it does not reflect the social value of the documentary process.

To combine the value of both definitions, we can link them together as follows:

A document is a trace for the interpretation of a past event in accordance with a reading contract. In most cases, this trace has been recorded on an appropriate medium for easy physical handling, transport across space and preservation over time.

7. Treading New Ground

If we pursue this line of reasoning, the proliferation of documents and of whole new genres of documents in all kinds of fields will reveal a feverish relationship to our past, a kind of existential questioning of our present, and a relationship to a future that is constantly being redefined before our eyes.

Digital documents, in particular, now form in their entirety a huge set of signals in an ever-widening stream. Everyone publishing on the web

can feel that he is an author. Indexing is no longer the preserve of librarians, and metadata have become indispensable to web ‘navigation.’ The claims of journalists are challenged by a multitude of internet sources. And the document as unitary artifact has been replaced by a set of scattered fragments that can be arranged and rearranged on the fly. These fragments form the leading edge of a hyperlinked network called the internet, and the traces left by its users now are used to optimize the system’s responses and form ever new corporate strategies. All of the various components of the document are hereby pounded into new shapes by forces moving along multiple dimensions, and the very idea of a document is itself made problematic.

Among the new professions that have emerged with the web, that of “information architect” perhaps best illustrates the new documentary state of affairs which now faces us. Information architects emphasize what they call the ‘user experience,’¹³ a notion which, when applied to information, has a strong kinship with the notion of reading contract as discussed above. However it lays greater stress on two important aspects. First is the active character of the reader—who is often also a writer. And second is the emotional aspect, drawn attention to in the use of the term ‘experience’ rather than ‘contract’. This latter aspect is critical in an environment of accelerating document overload, where it is essential to capture the fleeting attention of the reader.

The user experience flows through a variety of devices (computers, smartphones, tablets) and applications (search engines, social networks, blogs, tweets). The user is immersed in an environment where information-carrying objects communicate with one another and spawn new information-carrying objects, so that the reader must work hard to maintain consistency in his informational experience and to avoid cacophony and loss of meaning.

These trends are nicely summarized by Resmini and Rossati in their “manifesto” for a ubiquitous information architecture, with which I conclude this essay.¹⁴ To see the relationship with what I have just described, the reader needs merely to replace ‘information architecture’ by ‘document’ and ‘user experience’ by ‘reading contract’.¹⁵

Jean-Michel Salaün

Ecole Normale Supérieure de Lyon

APPENDIX:
PERVASIVE INFORMATION ARCHITECTURE—
DESIGNING USER EXPERIENCES. A MANIFESTO BY
ANDREA RESMINI AND LUCA ROSATI

Cyberspace is not a place you go to but rather a layer tightly integrated into the world around us.

—Institute for the Future

1. Information Architectures Become Ecosystems

When different media and different contexts are tightly intertwined, no artifact can stand as a single isolated entity. Every single artifact becomes an element in a larger ecosystem. All these artifacts have multiple links or relationships with each other and have to be designed as part of one single seamless user experience process.

2. Users Become Intermediaries

Users are now contributing participants in these ecosystems and actively produce new content or remediate existing content by way of mash-ups, commentary, or critique. The traditional distinction between authors and readers, or between producers and consumers, becomes weakened to the point of being useless and void of meaning.

All build new relationships and meanings by means of mash-ups, aggregators, and social networking tools, and all agents contribute content through the crowdsourcing possibilities leveraged by the Web via wikis, blogs, and other participatory tools, and mobile devices.

3. Static Becomes Dynamic

On one hand, these architectures aggregate and mash-up content which physically may reside elsewhere and which might have been released for completely different purposes. On the other hand, the active role played by intermediaries makes them perpetually unfinished, perpetually changing, and perpetually open to further refinement and manipulation.

4. Dynamic Becomes Hybrid

These new architectures embrace different domains (physical, digital, and hybrid), different types of entities (data, physical items, and people)

and different media. As much as the boundaries separating producers and consumers become weakened, so do those between different media and genres. All experiences are cross-media bridge-experiences across a breadth of different environments.

5. Horizontal Prevails Over Vertical

In these new architectures correlation between elements becomes the predominant characteristic, at the expense of traditional top-down hierarchies. In open and ever-changing architectures hierarchical models are difficult to maintain and support, as intermediaries push towards spontaneity, ephemeral or temporary structures of meaning, and constant change.

6. Product Design Becomes Experience Design

When every single artifact, be it content, product, or service, is a part of a larger ecosystem, focus shifts from how to design single items to how to design experiences spanning processes. Everyday shopping does not concern itself with the retail shop only, but configures an experience process which might start on traditional media with a television commercial or newspaper advertisement, might continue on the Web with a search for reviews or for location of the nearest convenience store, might proceed to the shop to finalize a purchase, and finally returns to the Web for assistance, updates, customization, and networking with other people or devices.

7. Experiences Become Cross-Media Experiences

Experiences bridge multiple connected media and environments into ubiquitous ecologies. A single unitary process where all parts contribute to the final, seamless user experience.

NOTES

1. ATILF (Analyse et traitement informatique de la langue française).
2. Paul Otlet, *Traité de documentation: le livre sur le livre. Théorie et pratique* (Bruxelles: Editions Mundaneum, 1934), 9. For a good review of definitions of 'document', see: Niels Windfeld Lund and Roswitha Skare, "Document theory," *Encyclopedia of Library and Information Sciences* (3rd edition) (New York: Taylor and Francis, 2010), 1632–39.

3. Suzanne Briet, *Qu'est-ce que la documentation?* (Paris: Éditions documentaires industrielles et techniques, 1951), 7. English translation by R.E. Day, L. Martinet and H. G.B. Angheliescu, *What is Documentation?* (Lanham, MD: Scarecrow Press, 2006).
4. <http://lccjti.ca>
5. Roger T. Pédaque, *Le Document à la lumière du numérique* (C&F, 2006). The texts are also available at [http:// archivesic.ccsd.cnrs.fr](http://archivesic.ccsd.cnrs.fr).
6. For a discussion of these three dimensions in English, see: Roger T. Pédaque, *Document: Form, Sign and Medium, as Reformulated for Electronic Documents*, translated by Niels Lund, 2003. http://archivesic.ccsd.cnrs.fr/sic_00000594/
7. It is a matter of neuroanthropology: see Stanislas Dehaene, *Reading in the Brain. The Science and Evolution of a Human Invention* (New York: Penguin, 2009).
8. For an overview, see Karl Canvat, "Pragmatique de la lecture: le cadrage générique," *Atelier de théorie littéraire: Genres et pragmatique de la lecture*, 2007, <http://fabula.org>.
9. *Lectures industrielles*, 2008 (<http://www.arsindustrialis.org/node/2879>).
10. Stanislas Dehaene, op. cit. Michel Serres, *Les Nouvelles Technologies: révolution culturelle et cognitive*, INRIA lecture of December 11, 2007 at <http://interstices.info>.
11. Pédaque, op. cit., 78.
12. Quoted by Michael Buckland, "What Is a 'Document'?" *Journal of the American Society of Information Science* 48, no. 9 (Sept. 1997): 804–07.
13. Garrett, Jesse James: *The Elements of User Experience: User-Centered Design for the Web* (Berkeley: Peachpit Press, 2002).
14. Resmini, Andrea, and Luca Rosati: *Pervasive Information Architecture: Designing Cross-Channel User Experiences* (Burlington MA: Morgan Kaufmann, 2011). See <http://pervasiveia.com/book/manifesto>.
15. In accordance with its subject matter, this document is being regularly reviewed. An initial French version was published in *E-dossiers de l'audiovisuel: Sciences humaines et sociales et patrimoine numérique*, INA-Sup, 2012, <http://www.ina-sup.com/node/2832>. This was already an abridged and revised version of Chapter 2 of the book by the same author: *Vu lu su: les architectes de l'information face à l'oligopole du web*, Paris: La Découverte (Cahiers Libres, 2012). It is also subjected to a comprehensive review before each season of courses, and thus reflects interactions with students. This version was translated into English by Christian Allegre, to whom many thanks are due.