UC Berkeley

Title
What is a "document"?

Permalink
https://escholarship.org/uc/item/8f4750f1

Journal
Journal of the Association for Information Science and Technology, 48(9)

ISSN
2330-1635

Author
Buckland, Michael K

Publication Date
1997-09-01

DOI
10.1002/(sici)1097-4571(199709)48:9<804::aid-asi5>3.0.co;2-v

Peer reviewed

What is a "document"?

Michael K. Buckland
School of Information Management & Systems, University of California, Berkeley, CA 94720-4600.

Abstract: Ordinarily the word "document" denotes a textual record. Increasingly sophisticated attempts
to provide access to the rapidly growing quantity of available documents raised questions about which
should be considered a "document". The answer is important for any definition of the scope of
Information Science. Paul Otlet and others developed a functional view of "document" and discussed
whether, for example, sculpture, museum objects, and live animals, could be considered "documents".
Suzanne Briet equated "document" with organized physical evidence. These ideas appear to resemble
notions of "material culture" in cultural anthropology and "object-as-sign" in semiotics. Others, especially
in the USA (e.g. Jesse Shera and Louis Shores) took a narrower view. New digital technology renews old
questions and also old confusions between medium, message, and meaning.

Introduction

What is a document? What could not be a document? Ordinarily information storage and retrieval
systems have been concerned with text and text-like records (e.g. names, numbers, and alphanumeric
codes). The present interest in "multimedia" reminds us that not all phenomena of interest in information
science are textual or textlike. We may need to deal with any phenomena that someone may wish to
observe: events, processes, images, and objects as well as texts.

This paper reconstructs and comments on the development of thought on this topic with an
emphasis on the ideas of continental European documentalists in the first half of this century. If
"documentation" (a term that included information storage and retrieval systems) is what you do to or
with documents, how far could you push the meaning of "document" and what were the limits to
"documentation"? The work of European pioneers such as Paul Otlet and Suzanne Briet has received
renewed attention in recent years and has been related to discussion of physical forms of "information"
(e.g. "information-as-thing" (Buckland 1991a, 1991b)). These issues are important because mechanical
information systems can only operate on physical representations of "information". This background is
relevant to the clarification of the nature and scope of information systems.

From document to "documentation"

In the late 19th century there was increasing concern with the rapid increase in the number of
publications, especially of scientific and technical literature. Continued effectiveness in the creation,
dissemination, utilization of recorded knowledge was seen as a needing new techniques for managing the
growing literature.

The "managing" that was needed had several aspects. Efficient and reliable techniques were
needed for collecting, preserving, organizing (arranging), representing (describing), selecting (retrieving),
reproducing (copying), and disseminating documents. The traditional term for this activity was
"bibliography". However, "bibliography" was not entirely satisfactory for two reasons: (i) It was felt that

something more than traditional "bibliography" was needed, e.g. techniques for reproducing documents; and (ii) "Bibliography" also had other well-established meanings, especially historical (or analytical) bibliography which is concerned with traditional techniques of book-production.

Early in the 20th century the word "documentation" was increasingly adopted in Europe instead of "bibliography" to denote the set of techniques needed to manage this explosion of documents. Woledge (1983) provides a detailed account of the evolving usage of "documentation" and related words in English, French and German. From about 1920 "documentation" was increasingly accepted as a general term to encompass bibliography, scholarly information services ("wissenschaftliche Aufklärung (Auskunft)"), records management, and archival work. (Donker Duyvis 1959. See also Björkbom 1959; Godet 1938).

There are numerous writings on the definition, scope, and nature of "documentation", much of it concerned with the relationships between documentation, bibliography, and librarianship. Unfortunately, many of this literature, like much of the later discussion of information science and librarianship, is undermined by the authors' attempts to create or amplify distinctions where the differences are not really fundamental but, rather, a matter of emphasis.
Loosjes (1962, pp. 1-8) explained documentation in historical terms: Systematic access to written texts, he wrote, became more difficult after the invention of printing resulted in the proliferation of texts; scholars were increasingly obliged to delegate tasks to specialists; assembling and maintaining collections was the field of librarianship; bibliography was concerned with the descriptions of documents; the delegated task of creating access for scholars to the topical contents of documents, especially of parts within printed documents and without limitation to particular collections, was documentation.

After about 1950 more elaborate terminology, such as "information science", "information storage and retrieval", and "information management", increasing replaced the word "documentation".

From documentation back to "document"

The problems created by the increase in printed documents led to development of the techniques of documentation. However, the rise of documentation led, in turn, to a new and intriguing question that received little direct attention then or since.

Documentation was a set of techniques developed to manage significant (or potentially significant) documents, meaning, in practice, printed texts. But there was (and is) no theoretical reason why documentation should be limited to *texts*, let alone *printed* texts. There are many other kinds of signifying objects in addition to printed texts. And if documentation can deal with texts that are not printed, could it not also deal with documents that are *not* texts at all? How extensively could documentation be applied? Stated differently, if the term "document" were used in a specialized meaning as the technical term to denote the objects to which the techniques of documentation could be applied, how far could the scope of documentation be extended. What could (or could not) be a document? The question was, however, rarely formulated in these terms.

An early development was to extend the notion of document beyond written texts, a usage to be found in major English and French dictionaries. (For historical background on "document" see also Sagredo Fernández & Izquierdo Arroyo (1982)). "Any expression of human thought" was a frequently used definition of "document" among documentalists. In the USA, the phrases "the graphic record" and "the generic book" were widely used. This was convenient for extending the scope of the field to include pictures and other graphic and audio-visual materials. Paul Otlet (1868-1944), is known for his observation that documents could be three dimensional, which enabled the inclusion of sculpture. From 1928, museum objects were likely to be included by documentalists within definitions of "document" (e.g. Dupuy-Briet 1933).

The overwhelming practical concern of documentalists was with printed documents, so the question of how far the definition of "document" could be extended received little direct attention.

Nevertheless, the occasional thoughtful writer would touch on the topic, perhaps because interested in some novel form of signifying object, such as educational toys, or because of a desire to generalize.

Paul Otlet: Objects as documents

Otlet extended the definition of "document" half-way through his *Traité de documentation* of 1934. Graphic and written records are representations of ideas or of objects, he wrote, but the *objects themselves* can be regarded as "documents" if you are informed by observation of them. As examples of such "documents" Otlet cites natural objects, artifacts, objects bearing traces of human activity (such as archaeological finds), explanatory models, educational games, and works of art (Otlet 1934, p. 217; also Otlet 1990, pp. 153 & 197, and Izquierdo Arroyo 1995).

In 1935 Walter Schuermeyer wrote: "Nowadays one understands as a document any material basis for extending our knowledge which is available for study or comparison." ("Man versteht heute unter einem Dokument jede materielle Unterlage zur Erweiterung unserer Kenntnisse, die einem Studium oder Vergleich zugaenglich ist." Schuermeyer 1935, p. 537).

Similarly, the International Institute for Intellectual Cooperation, an agency of the League of Nations, developed, in collaboration with Union Français des Organismes de Documentation, technical definitions of "document" and related technical terms in English, French and German versions and adopted:

> "Document : Toute base de connaissance, fixée matériellement, susceptible d'être utilisée pour consultation, étude ou preuve. Exemples: manuscrits, imprimés, représentations graphiques ou figurés, objets de collections, etc...
>
> Document : Any source of information, in material form, capable of being used for reference or study or as an authority. Examples : manuscripts, printed matter, illustrations, diagrams, museum specimens, etc....
>
> Dokument : Dokument is jeder Gegenstand, der zur Belehrung, zum Studium oder sur Beweisfuehrung dienen kann, z.B. Handschriften, Drucke, graphische oder bildliche Darstellungen, usw...." (Anon. 1937: 234)

Suzanne Briet: Physical evidence as document

One individual, who had, for years, been involved in discussions of the nature of documentation and documents, addressed the extension of the meaning of "document" with unusual directness. Suzanne Briet (1894-1989), also known as Suzanne Dupuy and as Suzanne Dupuy-Briet was active as a librarian and documentalist from 1924 to 1954 (Lemaître & Roux-Fouillet 1989; Buckland 1995).

In 1951 Briet published a manifesto on the nature of documentation, *Qu'est-ce que la documentation*, which starts with the assertion that "A document is evidence in support of a fact." ("Un document est une preuve à l'appui d'un fait" (Briet, 1951, 7). She then elaborates: A document is "any physical or symbolic sign, preserved or recorded, intended to represent, to reconstruct, or to demonstrate a physical or conceptual phenomenon". ("Tout indice concret ou symbolique, conservé ou enregistré, aux fins de représenter, de reconstituer ou de prouver un phénomène ou physique ou intellectuel." p. 7.) The implication is that documentation should not be viewed as being concerned with texts but with access to evidence.

The antelope as document

Briet enumerates six objects and asks if each is a document.

Object	Document?
Star in sky	No
Photo of star	Yes
Stone in river	No
Stone in museum	Yes
Animal in wild	No
Animal in zoo	Yes

There is discussion of an antelope. An antelope running wild on the plains of Africa should not be considered a document, she rules. But if it were to be captured, taken to a zoo and made an object of study, it has been made into a document. It has become physical evidence being used by those who study it. Not only that, but scholarly articles written about the antelope are secondary documents, since the antelope itself is the primary document.

Briet's rules for determining when an object has become a document are not made clear. We infer, however, from her discussion that:

1. There is materiality: Physical objects and physical signs only;
2. There is intentionality: It is intended that the object be treated as evidence;
3. The objects have to be processed: They have to be made into documents; and, we think,
4. There is a phenomenological position: The object is perceived to be a document.

This situation is reminiscent of discussions of how an image is made art by framing it as art. Did Briet mean that just as "art" is made art by "framing" (i.e. treating) it as art, so an object becomes a "document" when it is treated as a document, i.e. as a physical or symbolic sign, preserved or recorded, intended to represent, to reconstruct, or to demonstrate a physical or conceptual phenomenon? The sources of these views are not made clear, though she does mention in this context her friend Raymond Bayer, a professor of philosophy at the Sorbonne, who specialized aesthetics and phenomenology.

Ron Day (1996) has suggested, very plausibly, that Briet's use of the word "indice" is important, that it is indexicality--the quality of having been placed in an organized, meaningful relationship with other evidence--that gives an object its documentary status.

Donker Duyvis: A spiritual dimension to documents

Frits Donker Duyvis (1894-1961), who succeded Paul Otlet as the central figure in the International Federation for Documentation, epitomized the modernist mentality of the documentalists in his dedication to the trinity of scientific management, standardization, and bibliographic control as complementary and mutually reinforcing bases for achieving progress (Anon., 1964). Yet Donker Duyvis was not a materialist. He adopted Otlet's view that a document was an expression of human thought, but he did so in terms of his interest in the work of Rudolf Steiner (1861-1925), founder of Anthroposophy, a spiritual movement based on the notion that there is a spiritual world comprehensible to pure thought and accessible only to the highest faculties of mental knowledge. As a result, Donker Duyvis was sensitive to what we might now call the cognitive aspects of the medium of the message. He wrote that:

> A document is the repository of an expressed thought. Consequently its contents have a spiritual character. The danger that blunt unification of the outer form exercises a repercussion on the contents in making the latter characterless and impersonal, is not illusory.... In standardizing the

form and layout of documents it is necessary to restrict this activity to that which does not affect the spiritual contents and which serves to remove a really irrational variety." (Donker Duyvis, 1942. Translation from Voorhoeve, 1964, 48)

Ranganathan: Micro-thought on a flat surface

The Indian theorist S. R. Ranganathan, usually so metaphysical, took a curiously narrow and pragmatic position on the definition of "document", resisting even the inclusion of audiovisual materials such as radio and television communications.

But they are not documents; because they are not records on materials fit for handling or preservation. Statues, pieces of china, and the material exhibits in a museum were mentioned because they convey thought expressed in some way. But none of these is a document, since it is not a record on a more or less flat surface. (Ranganathan, 1963).

Ranganathan's view of "document" as a synonym for "embodied micro thought" on paper "or other material, fit for physical handling, transport across space, and preservation through time" was adopted by the Indian Standards Institution (1963, 24), with a note explaining that the term "document" "is now extended in use to include any embodied thought, micro or macro and whether the physical embodiment is exclusive to one work or is shared by more than one work.

Others, also, took a limited view of what documents were. In the USA, two highly influential authors opted for a view of documents that was only an extension of textual records to include audiovisual communications. Louis Shores popularized the phrase "the generic book" (e.g. Shores, 1977) and Jesse H. Shera used "the graphic record" with much the same meaning (e.g. Shera, 1972). Shera was gratuitously dismissive of Briet's notion of documents as evidence.

Anthropology: Material culture

Otlet was explicit that his view of "document" included archaeological finds, traces of human activity, and other objects not intended as communication.

Collections of objects brought together for purposes of preservation, science and education are essentially documentary in character (Museums and Cabinets, collections of models, specimens and samples). These collections are created from items occurring in nature rather than being delineated or described in words; they are three dimensional documents." (Otlet, 1920. Translation from Otlet 1990, 197).

The notion of objects as documents resembles the notion of "material culture" among cultural anthropologists "for whom artifacts contributed important evidence in the documentation and interpretation of the American experience." (Ames 1985, ix) and in museology (e.g. Kaplan 1994; Pearce 1990).

Semiotics: "Text" and "object-as-sign"

Briet's ideas concerning the nature of a "document" invite discussion in relation to semiotics. In this context we note Dufrenne's discussion of the distinction between aesthetic objects and signifying objects:

The function of such [signifying] objects is not to subserve some action or to satisfy some need but to dispense knowledge. We can, of course, call all objects signifying in some sense. However, we must single out those objects which do more than signify merely in order to prepare us for some action and which are not used up merely in the fulfillment of the task. Scientific texts, catechisms, photograph albums, and, on a more modest scale, signposts are all signs whose

signification engages us in an activity only after having first furnished us with information." (Dufrenne 1973, 114).

We can observe that by the inclusion of museum and other "found" objects, Briet's "any physical or symbolic sign" appears to include both human signs and natural signs. Others developed the notion of "object-as-sign". Roland Barthes, for example, in discussing "the semantics of the object", wrote that objects

function as the vehicle of meaning: in other words, the object effectively serves some purpose, but it also serves to communicate information: we might sum it up by saying that there is always a meaning which overflows the object's use." (Barthes, 1988, 182).

We can note the widespread use of the word "text" to characterize patterns of social phenomena not made of words or numerals, but there seems to have been relatively little attention to the overlap between semiotics and information science. (See, however, the careful discussion by Warner 1990.)

Comments

One difference between the views of the documentalists discussed above and contemporary views is the emphasis that would now be placed on the social construction of meaning, on the viewer's perception of the significance and evidential character of documents. "Relevance", a central concept in information retrieval studies, is now generally considered to be situational and ascribed by the viewer. In semiotic terminology,

...signs are never natural objects... The reason is simply that the property of being a sign is not a natural property that can be searched for and found, but a property that is given to objects, be they natural or artificial, through the kind of use that is made of them. Both as objects and as means, signs have to be treated as something invented, and in this sense they are correlated to actions." (Sebeok 1994, v. 1, p. 18).

Briet's notion of documents as evidence can occur in at least two ways. One purpose of information systems is to store and maintain access to whatever evidence has been cited as evidence of some assertion. Another approach is for the person in a position to organize artefacts, samples, specimens, texts, or other objects to consider what it could tell one about the world that produced it, and then, having developed some theory of its significance to place the object *in evidence*, to offer it as evidence by the way it is arranged, indexed or presented. In this manner information systems can be used not only in finding material that already is in evidence, but also in arranging material so that someone may be able to make use of it as (new) evidence for some purpose. (Wilson 1995).

The evolving notion of "document" among Otlet, Briet, Schürmeyer, and the other documentalists increasingly emphasized whatever functioned as a document rather than traditional physical forms of documents. The shift to digital technology would seem to make this distinction even more important. Levy's thoughtful analyses have shown that an emphasis on the technology of digital documents has impeded our understanding of digital documents as documents (e.g. Levy 1994). A conventional document, such as a mail message or a technical report, exists physically in digital technology as a string of bits, but so does everything else in a digital environment. In this sense, any distinctiveness of a document as a physical form is further diminished and discussion of "What is a digital document?" becomes even more problematic unless we remember the path of reasoning underlying the largely forgotten discussions of Otlet's objects and Briet's antelope.

Postscript: Documenting the antelope

Briet's discussion of an antelope as a document is quite specific: The antelope was from Africa; it was a newly discovered species; and it was brought to the Jardin des Plantes of the Muséum National

d'Histoire Naturelle in Paris. Her account reads as if she were referring to an actual antelope known to her. In 1947, not long before Briet's book appeared, the Muséum National d'Histoire Naturelle did announce the discovery of a new African antelope - *tragelaphus scriptus reidae*, a subspecies of bushbuck - but there is no indication that a specimen was taken to Paris (Babault 1947). The documentation of antelopes reveals that very few new species were discovered during Briet's lifetime and documentary evidence of the Briet's antelope has eluded us. Appropriately, the word "antelope" itself, we found, is thought by some to derive from the Ethiopian word for the elusive unicorn.

Acknowledgements

References

Ames, K. L. et al. (1985). *Material culture: a research guide*, ed. by T. J. Schlereth. Lawrence, KS: University Press of Kansas.

Anon. (1937). La terminologie de la documentation. *Coopération Intellectuelle. 77,* 228-240.

Anon. (1964). *F. Donker Duyvis: His life and work*. The Hague: Netherlands Institute for Documentation and Filing. (NIDER publ. ser. 2, no. 45). 39-50.

Babault, G. (1947). Description d'une nouvelle sous-espèce du genre Tragelaphus (Mammifère ongulé). Tragelaphus scriptus reidae. *Bulletin du Muséum National d'Histoire Naturelle 2nd ser., 19*, 379-380.

Björkbom, C. (1959). History of the word documentation within the FID. *Revue de Documentation, 26*, 68-69.

Briet, S. (1951). *Qu'est-ce que la documentation*. Paris: EDIT.

Buckland, M. K. (1991a). *Information and Information Systems*. New York: Greenwood; Praeger.

Buckland, M. K. (1991b). Information as thing. *Journal of the American Society of Information Science* 42:5 (June 1991): 351-360.

Buckland, M. K. (1991c). Information retrieval of more than text. *Journal of the American Society for Information Science, 42*, 586-588.

Buckland, M. K. (1995). The centenary of `Madame Documentation': Suzanne Briet, 1894-1989. *Journal of the American Society for Information Science, 42*, 586-588.

Buckland, M.K. & R. Day. The semiotics of "document". In the proceedings of the Fifth Congress of the International Association for Semiotic Studies, Berkeley, 1994. Berlin: Mouton de Gruyter, forthcoming.

Donker Duyvis, F. (1942). *Normalisatie op het gebied der documentatie*. [Standardization in the domain of documentation]. (NIDER publ. 214). The Hague, Netherlands: NIDER.

Donker Duyvis, F. (1959). Die Enstehung des Wortes `Dokumentation' im Namen des FID. *Revue de Documentation, 26*, 15-16.

Dufrenne, M. (1973). *The phenomenology of aesthetic experience*. Evanston, IL: Northwestern UP.

Dupuy-Briet, S. (1933). Rapport présenté à la Commission de terminologie. In *International Institute for Documentation. XIIe Conférence. Rapport*. Bruxelles, 1933 (pp. 187-192). IID publication 172a. Brussels:IID.

Frank, P. R. (1978). *Von der systematischen Bibliographie zur Dokumentation*. (Wege der Forschung 144). Darmstadt: Wissenschafliche Buchgesellschaft.

Godet, M. (1938). Documentation, bibliothèques et bibliographie: Essai de définition de leurs caractères et de leur rapports. *IID Communicationes, V*, Fasc. 1, 15-18.

Indian Standards Institute. (1963). *Indian standard glossary of classification terms. IS : 2550 - 1963.* New Delhi: Indian Standards Institute.

Izquierdo Arroyo, J. M. (1995). *La organizacion documental del conocimiento*. Madrid: Tecnidoc.

Kaplan, F. E. S. (Ed.) (1994). *Museums and the making of "ourselves": The role of objects in national identity*. London: Leicester University Press.

Lemaître, R., & Roux-Fouillet, P. (1989). Suzanne Briet (1894-1989). *Bulletin d'Informations de l'Association de Bibliothecaires Français, 144*, 55-56.

Levy, D. M. (1994). Fixed or fluid? Document stability and new media. In *Eurpean Conference on Hypertext Technology 1994 Proceedings*. (Pp. 24-31). New York: Association for Computing Machinery.

Loosjes, T. P. (1962, 1967). *Dokumentation wissenschaftlicher Literatur*. Munich: BLV Verlagsgesellschaft, 1962. Chap. 1. Was ist Dokumentation? English translation: On *Documentation of Scientific Literature*. London: Butterworths, 1967.

Otlet, P. (1920). *L'organisation internationale de la bibliographie et de la documentation*. IIB Publ. 128. Brussels: Institut International de Bibliographie. Translation in Otlet (1990) pp. 173-203.

Otlet, P. (1934 [1989]). *Traité de documentation*. Brussels: Editiones Mundaneum. Reprinted 1989, Liège: Centre de Lecture Publique de la Communauté Française.

Otlet, P. (1990). *International organization and dissemination of knowledge: Selected essays*. (FID 684). Amsterdam: Elsevier.

Pearce, S. M. (Ed.) (1990). *Objects of knowledge*. (New research in museum studies, 1). London: Athlone Press.

Ranganathan, S. R., ed. (1963). *Documentation and its facets*. London: Asia Publishing House.

Sagredo Fernández, F. & Izquierdo Arroyo, J. M. (1982). Reflexiones sobre "Documento": Palabra / objecto. *Boletin Millares Carlo, 3*, 161-197.

Schuermeyer, W. (1935). Aufgaben und Methoden der Dokumentation. *Zentralblatt für Bibliothekswesen, 52*, 533-543. Repr. in Frank 1978, pp. 385-397.

Sebeok, T. A., ed. (1994). *Encyclopedic dictionary of semiotics*. 2nd ed. Berlin: Mouton de Gruyter.

Shera, J. H. (1972). *The foundations of education for librarianship*. New York: Becker and Hayes.

Shores, L. (1977). *The generic book: What is it and how it works*. Norman, OK: Library-College Associates.

Voorhoeve, N. A. J. (1964). F. Donker Duyvis and standardization. In: *F. Donker Duyvis: His life and work*. The Hague: Netherlands Institute for Documentation and Filing. (NIDER publ. ser. 2, no. 45). 39-50.

Warner, J. (1990). Semiotics, information science, documents and computers. *Journal of Documentation, 46*, 16-32.

Woledge, G. (1983). `Bibliography' and `documentation': words and ideas. *Journal of Documentation, 39*, 266-279.