Document Theory: An Introduction.

Preprint, Nov 7, 2013, of presentation at the Summer School on Records, Archives and Memory Studies, University of Zadar, Department of Information Sciences, Zadar, Croatia, 6 to 10 May, 2013 to be published in *Summer School on Records, Archives and Memory Studies: Proceedings* / edited by Anne J. Gilliland, Marijana Tomić and Mirna Willer. Zadar: University of Zadar, (forthcoming, 2014 or 2015). Published version may differ slightly.

Michael Buckland
School of Information, University of California, Berkeley, CA 94720, USA

Abstract
Writing, printing, telecommunications, and copying enabled the rise of the “information society” (more accurately “document society”) characterized by the division of labor. *Document* (verb) means to make evident. A *document* (noun) is something from which you learn, especially a text. Documents have a phenomenological aspect; employ cultural codes; form media types; and use physical media. The management of documents led to a more inclusive definition, including Briet’s antelope. Documents have technical, social, and mental aspects. Factual assertions require context. Documents are used to shape our lives and culture. Conventionally documents are made as documents; but objects can also be made into documents; or simply regarded as documents. New forms of document require new forms of bibliography. Current trends lead to the recording, the representation, and the analysis of everything, simultaneously.

**Keywords:** Document, Information, Society, Technology.

Many fields of study include consideration of documents. Paleographers examine handwriting on documents; archivists worry about the provenance and authenticity of documents; bibliographers are concerned to describe the content and physical construction of individual documents; and so on. *Document theory*, which overlaps with all of these, starts with the notion of a *document* as its point of departure. It is, if you will, a document-centric view of the universe and as the view expands, it progressively connects with the other fields of study in which documents are important but not the central focus.

**Information society, document society**

It has been fashionable to speak of a new or emerging “information society” but every society is an “information society” and always was because all communities, both human and animal, are formed by communication, interaction, and collaboration. All depend on communication, on information. There cannot be a “non-information society”!

Members of communities, humans and some other animals, communicate through gesture, language, and the use of material objects to signify something. Social interactions and social control are increasingly indirect and through documents. We depend more and more on documents. So “document society” would be more accurate than “information society”. Patrick Wilson wrote of this trend as “second-hand knowledge.”

\footnote{Wilson, Patrick. Second-hand knowledge: An inquiry into cognitive authority. Westport, CT: Greenwood, 1983.} It can be seen as having two aspects. There is an ever increasing division of labor, which makes us more dependent on others and which requires ever-increasing coordination through communication, and there is more and more dependence on documents as the means for this
communication. In a significant sense, therefore, the phenomenon of the document society (also known as the information society) is both a by-product of the division of labor and a necessary part of the infrastructure than allows and enables that division.

**Terminology**

*Document*, as a verb, means to make evident, to provide an explanation.

*Document*, as a noun, was, historically, something you learned from, including a lesson, a lecture, or an example. Gradually, *document* came increasingly to mean a text, but retaining a sense of evidence.

During the twentieth century the word *information* became increasingly popular and was used with many different meanings which largely fall into one of three categories:

-- Information-as-knowledge, meaning the knowledge imparted through communication;
-- Information-as-process, the process of becoming informed; and
-- Information-as-thing, denoting bits, bytes, books, and other physical media. This is the commonest use of the word ‘information’ and can include any material thing or physical action perceived as signifying. In this third sense, ‘information’ becomes a synonym for broad view of ‘document.’

**A concise history of documents long-term**

Among prehistoric humans and among at least some animals communication was through gestures, language, and the display of objects. Four major developments have changed that situation:

1. **Writing.** Writing mediates speech, a performative, ephemeral form of expression, by making it endure, thereby diminishing the effect of time. The temporal stability of recorded speech facilitates continuity and prolongs its influence and so enhances its effect. One consequence was the rise of records to augment or replace human memory. There is a large literature on the transition from oral to literate cultures.

2. **Printing.** Printing constitutes an extreme multiplication of writing, making it and speech that it records more productive. There is also an extensive literature on the impact of printing, especially the impact of printing with moveable type in the fifteenth century Europe which facilitated the Reformation and the rise of modern science.

3. **Telecommunications.** Until the nineteenth century telecommunications was, with few exceptions, a matter of someone going on foot, by horse, or by boat to deliver good news or bad. Then a series of innovations, notably the semaphore, telegraph, telephone, radio, and television led to the present digital environment. Telecommunications diminish the effect of distance and delay. Much like printing, telecommunication facilitates coordination, cohesion, and control and is the subject of a large literature.

4. **Document copying.** Transcribing texts is as old as writing, but techniques for generating rapid, reliable, economical copies of documents is essentially a twentieth-century development. There were three really important techniques: Photostat, microfilm, and electrostatic copying (commonly known as xerography). In striking contrast to writing, printing, and telecommunications, there has been little historical or social commentary on the impact of modern copying technology.

**Three copying technologies**

---

In the eighteenth century manuscripts were copied by “letter press”: A thin moist sheet of paper was pressed against the original so that some of the ink of the original would transfer into the moist sheet. During the nineteenth century photography was used occasionally to copy documents. (The numerous forms of duplicating involve the creation of a new original and are more properly regarded as small-run printing).

Photostat, direct-projection photography on to sensitized paper without an intermediate negative film, was pioneered by René Graffin of the Institut Catholique in Paris to facilitate his editing of early Christian writings in Syriac. The image was negative: white writing in a black ground. Left-to-right reversal was corrected by using a mirror. A few photostat cameras were built for European libraries but there was little impact until the Rectigraph Camera Company began marketing photostat equipment in the USA in 1910. The speed, accuracy, and efficiency of photostats for both text and images compared with manual transcription or typewritten copies were quickly recognized. The photostat process was widely adopted and became the copying process of choice at least until the late 1930s. Microfilm was famously used by Dagron in the siege of Paris in 1870, but widespread use was delayed until compact precision cameras, standard film speeds, and 35 mm safety film all became available during the 1930s when banks, newspapers, libraries and other organizations adopted microfilm and its variants (microfiche, microcard) on a large scale.

Electrostatic copying, better known as xerography, was developed to replace photostats, became available during the 1960s, and remains the technology of choice for copying and for printing digital documents.

Copying techniques need to make legible copies from very varied originals so copying was not in practice separable from image enhancement. A normal photograph of a medieval vellum manuscript using visible light would copy what one saw, but photography using ultraviolet light would also reveal erased text. From a technical perspective it is best to consider document copying jointly with image enhancement.

Perception of documents

Documents are concerned with evidence and evidence implies facts. Meaning, however, is constructed in the mind of the observer, even where facts are concerned. Paul Otlet considered that most authors were too wordy, that texts tended to be duplicative, and the print-on-paper bound book (codex) was an inefficient medium. Expressing an idea does not naturally fit line length or page size. He wanted to extract facts and ideas documents and to design a new and better form of “the book.” Otlet’s principal explanation of his ideas, his Traité de documentation, was published in 1934.

---


The very next year, Ludwik Fleck, publishing his explanation of why concise encyclopedias were fundamentally wrong in his *Genesis and development of the scientific fact*. Fleck argued that extreme summarization became misleading because too much of the contextual explanation was omitted. Further, he argued, it was fundamentally inadequate to think only in terms of a *fact* and a *reader*, because how a narrative statement of fact would be understood by a writer would depend on prior knowledge and the cultural context of a mindset (*Denkstil*) within a community of thinking (*Denkkollektiv*) of the writer (Sady 2012). Similarly, the understanding of the same fact by a reader would depend of the mindset and cultural context of the reader. As one simple illustration we cite Paracelsus, the colorful physician and scientist, struggling to express his ideas during the transition from alchemy to early modern chemistry. He lacked the concepts and terminology that were later to emerge. As a result we cannot understand much of what he wrote and he would not have understood present-day texts even on the topics in which he specialized. Facticity, then, is culturally situated, so documents have an evidentiary component that cannot be understood in only technological terms.

**Documents and technology**

Any document has multiple aspects:

1. **Phenomenological aspect.** There is a phenomenological aspect to documents. So long as documents are objects perceived as signifying something, the status of being a document is not inherent (essential) but attributed to an object. Meaning is always constructed by a viewer. We can view the properties of a document as having three aspects:

2. **Cultural codes.** All forms of communicative expression depend on some shared understandings, which can be thought of as language in a broad sense;

3. **Media types.** Different types of expression have evolved: texts, images, numbers, diagrams, art, music, dance, . . .

4. **Physical media.** Clay tablets, paper, film, analog magnetic tape, punch cards, digital media, and so on.

The status of being a document, therefore, is attributive (1) and every document has cultural (2), type (3), and physical (4) aspects. Genres are culturally and historically situated combinations. Being digital affects directly only the physical medium, but the consequences are extensive.

**The Documentation Movement**

In Brussels in 1895 Paul Otlet and Henri LaFontaine founded an International Institute for Bibliography to resolve the difficulties created by the ever-increasing number of documents. Later they later adopted *documentation*, rather than bibliography, as the term of choice for their activities. Their mission had important consequences with respect to design and scope. The standard technology, the printed, bound book had disadvantages: The format was inflexible; the divisions into lines and pages were irrelevant to the meaning, and the narratives, being both internally redundant and duplicative of other texts, were inefficient. Their answer, shared with Wilhelm Ostwald and H. G. Wells, was to prefer a synthesized, updateable, hypertextual encyclopedia on cards which, because it could record in one system everybody’s knowledge, was considered a “world brain.”

Documentation, then, was concerned with the management (selection, collection, arrangement, indexing, etc.) of documents and it was inevitable that a question would arise: With what kinds of document was documentation concerned? Printed texts were the primary concern, but since documents were of interest because they were evidence of something, then handwritten texts should also be included. And since diagrams, drawings, maps and photographs are used to describe or explain, images should not

---

be excluded. Once one accepts the notion of documents as objects from which one may learn, then there is no basis for limiting the scope to texts recorded on two-dimensional, flat surfaces. If a plans and maps can be documents, there is no reason to exclude relief maps and terrestrial globes. If a drawing can be a document, then why not a three-dimensional image (sculpture)? If a diagram, why not an educational or illustrative model or an educational toy? If three-dimensional objects are included, why not biological specimens, archaeological finds, and museum objects? Once three-dimensional objects are included, museum specimens and expressive sculpture cannot reasonably be excluded. If written language is included, then why not recorded spoken language or music? And if recorded speech and music are included, why not live performances?7

The compelling logic that since documents were, by definition, concerned with evidence, so too was bibliography and documentation, was advanced further by Suzanne Briet in her manifesto What is documentation?

“A document is a proof in support of a fact. . . .

Is a star a document? Is a pebble rolled by a torrent a document? Is a living animal a document? No. But the photographs and catalogues of stars, the stones in a museum of mineralogy, and the animals that are catalogued and shown in a zoo are documents.”8

Famously, she declared than a new species of antelope placed in a taxonomy and in a cage was a primary document and that descriptions of it in scientific articles, lectures, news reports, and encyclopedias were secondary documents.

After 1945 the documentation movement was largely forgotten until three factors led to a revival of interest in the 1990s. One was a new interest in the history of documentation and information science; another was inquiry into how digital documents differed from paper one; and, quite separately, legislation establishing a national library in Norway required the legal deposit of new media. Anticipating the need for additional expertise in handling new media, an Institute for Documentation Studies at the University of Tromsø initially headed by Niels W. Lund.9

**Lund’s Complementary Theory of Documents**

Niels W. Lund developed a “complementary” theory of documents, arguing that documents have three simultaneous, inseparable, and complementary aspects:

1. Technical and technological aspects of documents;
2. Social roles of documents; and
3. Mental: The intellectual and cognitive aspects of the relationship between an individual and a document.

---

“...one should view the document from three complementary angles: physical, social, and mental, in combination enabling a complete description. This does not mean that the document possesses each of these features to a some degree but that it can be viewed simultaneously as a physical, social, and mental phenomenon. From this perspective is how these dimensions interact with each other in different ways.”

The Literature on Document Theory

Much has been written on many different aspects of documents in bibliography, paleography, diplomatic, book design, publishing, archives, and so on, but the literature directly and simply about documents is relatively small. A convenient review of that literature has been provided by Lund in 2009 who stated that the literature is largely divided into two components: Technical, concerned with documents themselves; and social, concerned with the role of documents.

1. Technical. The professional theory of documents and documentation, concerned directly with documents themselves and their contents, is mainly technical and technological. The importance of documents means that there are well-established interests in them in a variety of fields, including:
- bibliography concerned with the description of documents and selection of documents for any specific purpose.
- Information retrieval concerned with sorting and selection of records in relation to the problematic notion of relevance;
- Bibliometrics (citation analysis and the like) concerned with formal, quantitative relationships between documents and between people and documents; and
- Textual studies, including the interpretative examination of relationships between texts

These various specialties overlap with documentation. The overlap is to be welcomed because they bring additional shared energy and resources to bear. The differences are mainly a matter of emphasis and/or perspective. Libraries, museums, archives, and other engaged in cultural agendas through the suitable management of documents. Suzanne Briet characterized documentation as a “a new cultural technique” and as “a necessity for our time.”

2. Social. “General document theory”, is concerned with what documents do, or, more properly, what is done with documents. It is not only our needs that should concern us, but also the agendas of others. Governments of passports to indicate identity and to confirm citizenship in order to control our travel. Other examples are easy to find: schools use textbooks to guide our learning; religions use sacred texts to inspire beliefs; artists produce images to please us and to challenge us; merchants invest heavily in advertisements to influence what we buy; politicians make statements to seek votes and campaign donations; entertainers use varied media to amuse us and to generate income from us; individuals use messages (letters, e-mail, etc.) to communicate and social media to attract attention; museums use the selective presentation and interpretation of artifacts to explain the past; mass media constantly transmit programs to entertain, to influence us, and to satisfy advertisers; libraries provide access to selective collections of documents to facilitate our reading; and so on. This list could be extended indefinitely.

Anyone could easily make such a list. As the list builds up we see more and more of our lives included. The specific examples enumerated are less important than the cumulative evidence that documents are everywhere in our lives and they shape our society and our culture. The use of documents involves far more than fact-finding and problem solving. The complex totality of our lives and of our behavior is our culture. Tylor’s classic definition of culture included “knowledge, belief, art, morals, law, custom, and

---

any other capabilities and habits acquired by man as a member of society.” Documents, then, are used to shape our culture.

The literature on the social theory of documents has been largely outside of Library and Information Science. Components include the histoire du livre tradition and work on the “social life of documents” by J. S. Brown, Paul Duguid, and others. Michel Foucault drew attention to the use of documents in the construction of reality and, thereby, in the shaping of power relations. Karl Manheim used the phrase “documentary meaning” to refer to the unintended meanings associated with documents. JoAnn Yates showed the impact of the evolving use of documentary forms in the management of large organizations. Donald McKenzie’s lecture “Bibliography and the sociology of texts (1986) was an eloquent plea for both increased attention to the social aspects of document creation and use and also for including a wider variety of document types in bibliography. Bernd Frohmann’s Deflating information (2004) has much to say about the use of documents.

Frontiers and agendas: Documents

The definition of document is still not settled. I find it helpful to identify three views:

1. The conventional, material view: The everyday, conventional view of documents is of graphic records, usually text, made on paper (or similar: clay tablets, microfilm, word processor files) that are material, local, transportable. On the fringes, some would argue, are terrestrial globes and sculptures. These objects are made as documents.

2. A functional view: Almost anything can be made to serve as a document, to signify something, to be held up as constituting evidence of some sort. Models, educational toys, natural history collections, and archaeological traces can be considered in this category. Briet’s famous antelope, positioned in a taxonomy and in a cage, was made to serve as a document. This view echoes her assertion that bibliography (and documentation) is properly considered to be concerned with access to evidence not just records. These are objects made into documents.

3. A semiotic view: The two previous views emphasize the creation of documents and they both seem inadequate because anything could be considered as a document if it is regarded as evidence of something regardless of what, if anything, the creator (if any) of that object intended.

These three views – made as, made into, and considered as – are progressively more inclusive.

Frontiers and agendas: Bibliography

Bibliography is concerned with the twin tasks of describing and selecting the best documentary means for some purpose. We may choose to limit ourselves to a single kind of document, such as printed books, but in principle bibliography is not and should not be so restricted. We can, therefore, spell out the scope of bibliography by enumerating the conditions that must be satisfied if one is to use a documentary means to some purpose. What follows is one such enumeration.

For a document to be used:

References:

0. Creation: It must exist;
1. Discovery: We need to know of its existence;
2. Location: We need to find a copy;
3. Permission: We may need permission to use it. There may be legal constraints.
4. Condition: Is it in a fit state to use? Is it too deteriorated and/or too obsolete to be worth using?
5. Interoperable: Is it standardized enough to be usable? Digital or microform materials may require unavailable equipment.
6. Description: Is it clear enough what it represents?
7. Trust: Are we confident enough of the origin, lineage, version, and error rate?

These requirements are different in kind and call for different sorts of remedy, some more feasible and/or more affordable than others. All need to be resolved for satisfactory document use.15

Frontiers and agendas: The third literature

It is noticeable that Lund identified three aspects of document theory but only two literatures. His category “professional document theory” literature corresponds, more or less to the “technical aspect” of documents and what he described as “social document theory” literature approximates the his “social aspect” of documents. How is this to be explained? It could, of course, be simply a matter of differing definitions, but it seems not. The inference is that there should be, a third literature corresponding to his “mental” aspect of documents concerned with the individual’s cognitive and intellectual engagement with documents in addition to the massive literature on information seeking and searching. Is there? If so, where is it and what does it include? It is not that there is no such literature but rather that a coherent overview in relation to documents is needed.

A literature addressing the cognitive / intellectual aspect of documents would presumably address the individual’s mental relationship with documents, whether as creator or perceiver. In general terms, one might reasonably expect to look to studies of rhetoric for the former and of semiotics for the latter. More specialized fields where one might look include educational psychology, hermeneutics, and reading theory. These fields are very little present in the literature of documentation, at least by those names.

A basic issue is that what anyone is likely to learn from a document will be heavily influenced by what that person already knows. Allan Konrad has identified a massive failure of LIS literature to take that aspect into account.16 To make progress in this matter will require a sharp focus on why and how documents are used.

Frontiers and agendas: Foundations

Culture and society evolve through communication and collaboration. But mostly we cannot communicate directly and we cannot collaborate in shared work environment for multiple very practical reasons. The other person of interest may be distant in space or time – and may even be dead. The best we can do is to have recourse to documents. What is that other person known to have said, done, or written? Their documents, both by them and about them, incorporate their work and their ideas, much as technology incorporates the labor of past inventors. Viewed this way, documents have become the glue

that enables societies to cohere. Documents have increasingly become the means for monitoring, influencing, and negotiating relationships with others. We live in a document society.

Much has been made of the transition from an oral to a literate culture and, how, for example, the ability to record what we need to remember has weakened our cerebral memories. As stated this is a gross simplification. First, the emphasis on orality disregards the important communicative roles of dance, music, and ritual. Second, the effect was additive: literacy was added to orality.

There is far more to the use of documents than literacy because not all the records that affect us are humanly readable. They are neither read nor acted upon by humans, at least not directly. For examples, commercial and transportation are now dominated by printed bar codes. We may see them and we know what they are, but we cannot ourselves read them. In the emerging digital environment of bar codes, sensors, and remote databases, increasingly the documents that shape our lives are not humanly readable. Often, they are not visible to a human eye. So now that we are in a document society, wherein a sense of documentary processes (“documentality”) is added to literacy and orality.

A concise future of documents long-term

We have offered, above, a concise history of documents since prehistoric times. What do we find when we extrapolate the past into the future using the same components and assuming continuing improvements in technology?

1. Writing, a means for the recording of speech is moving steadily towards the recording of everything.

2. Printing, the multiplication of texts, is evolving into the representation of anything, especially now with 3D printers.

3. Telecommunications, in effect the transportation of documents, becomes, with sustained improvement, effectively pervasive simultaneous interaction.

4. Document copying, because it depends for versatility on the use of image enhancement, leads to more than the making of additional copies. The logical development of document copying is the analysis of documents, including data sets and visualization.

The simple extrapolation of past trends, then, leads to a document society characterized by ubiquitous recording, pervasive representations, simultaneous interaction regardless of geographical distance, and powerful analysis of the records resulting from that ubiquitous recording. It is not merely that new technology enables the rise of new document genres. It also allows disparate genres to be woven together much more completely than before, a new tapestry.

References


Biographical sketch
Michael Buckland is Emeritus Professor in the School of Information, University of California, Berkeley. Born and educated in England, he studied History at Oxford University and Librarianship at Sheffield University and has worked as a librarian in England and in the U.S.A. He has served as Dean of the School of Library and Information Studies at Berkeley, as Assistant Vice President for Library Plans and Policies for the University of California, and as President of the American Society for Information Science. He is interested in the planning of library services, the organization of knowledge, cultural heritage, and the history and theory of documentation. His publications include Information and information systems (1991), Redesigning library services (1992), and Emanuel Goldberg and his knowledge machine (2006).