Title
Document Theory

Author
Buckland, Michael

Peer reviewed

Document Theory

Michael Buckland
School of Information,
South Hall 102,
University of California,
Berkeley, CA 94720-4600

Abstract

Document theory examines the concept of a document and how it can serve with other concepts
to understand communication, documentation, information, and knowledge. Knowledge
Organization itself is in practice based on the arrangement of documents representing concepts
and knowledge. The word *document* commonly refers to a text or graphic record, but, in a
semiotic perspective, non-graphic objects can also be regarded as signifying and, therefore, as
documents. The steady increase in the variety and number of documents since prehistoric times
enables the development of communities, the division of labor, and reduction of the constraints
of space and time. Documents are related to data, facts, texts, works, information, knowledge,
signs, and other documents. Documents have physical (material), cognitive, and social aspects.

1. Introduction

Document theory is a field that examines both the concept of a document and how it can serve
with other concepts to understand better the complex areas of communication, documentation,
information, and knowledge.

1.1. Knowledge organization and document theory

Knowledge organization is concerned with describing, representing, organizing, discovery,
selection, and retrieval of concepts and of knowledge in a wide variety of contexts. As a practical
matter it does this through the manipulation of *representations* of concepts and of knowledge,
through documents. Knowledge Organization is, therefore, directly and centrally concerned with
documents. In consequence, understanding the nature of documents should be an important
concern in knowledge organization.

1.2. Etymology

Theory is a view, a perception, an understanding of something, and it is etymologically related to
theater. *Document theory*, therefore, should be regarded as a perspective on documents, their
nature, any aspect of them, and their role.

Document, like cognate words in English (e.g. *docent*, *docile*, *doctor*, *doctrine*) and other Western languages, derives from the root of the Latin verb *docere* which has a range of meanings centered on teaching, instructing, showing, telling, and demonstrating. The qualifying suffix *-mentum* denotes a non-abstract object. The English noun *document* was adopted via Old French from the Latin *documentum*, which denoted a lesson, proof, instance, or a specimen, and was increasingly used to refer to a written instrument, charter, or official paper. In modern times *document*, as a noun, came to mean almost exclusively a textual or graphic record on paper or, now, an electronic medium, but the sense of showing or instructing remained.

Using *document* as verb, *to document* something, denotes the creating of didactic or evidentiary records of some thing or some process: The creation of *phenomena* (perceptible things-for-us) representing a possibly imperceptible *noumenon* (thing-in-itself).

Documentation denotes either the process or the outcome of documenting. *Documentary*, an adjective, means having the character of a document, i.e. representing or explaining something, especially when using film as a medium. *Documentality* has been used by Frohmann (2011) to denote the properties and traces by which an object achieves its documentary role and by Ferraris for the character of traces and inscriptions in his theory of social ontology (Ferraris 2013; also Ferraris and Caffo 2014).

1.3. Scope

The study of documents has been heavily concerned with the many varieties of physical forms of documents, how they relate to each other and to their contexts, how they are or should be used, efforts to exercise control over them, and, especially, the development of new capabilities using new technologies: writing, paper, printing, photography, and, more recently, electronics. These important topics are addressed in bibliography, media studies, publishing, information technology, and elsewhere, and are beyond the scope of this article on document theory, which is concerned with ideas concerning the concept of a document.

2. History

After prehistoric times, improved technology allowed additions to the repertoire of dance, drawing, gesture, speech, and ritual. By recording, writing provides an alternative to speech. Printing and copying provided multiplicity of copies, and successive developments in telecommunications have allowed rapid transmission of messages. These developments had two consequences: the effects of time and distance were steadily mitigated; and, combined with the progressive division of labor in society, led to a dramatic increase in both the number of documents and in our dependence on them. Referring to an emerging *information society* is incorrect because all societies, including prehistoric hunters and gatherers, are constituted by and depend on collaboration and the sharing of information. It is the ever-increasing role and number of documents that is new and significant. Increasingly we live in a *document society* (Buckland 2017a) in which we depend more and more on recorded statements, on second-hand knowledge (Wilson 1983).

The rise of documents induced a need for a new line of technical development—known as bibliography, documentation, librarianship, information management, and by other names—to cope with the flood and to provide for discovery of and access to desired documents as and when needed. It was in this context that modern document theory arose. The increase in documents led to increased attention to bibliographic access. And as bibliography (later the term *documentation* was also used) is seen as being concerned with documents, it was inevitable that a question would arise concerning the scope and range of this field. To what range of objects can the process of documentation be applied? What objects can qualify as documents? So long as the evidentiary aspect is important, a narrowly material definition becomes unsatisfactory. Paul Otlet's *Traité de documentation* [Treatise on documentation] (Otlet 1934) starts by defining *document* broadly to include not only paper records, but also photographs, films, and statistical data (p. 43). Then, later on, he abruptly widens the scope further: Graphic and written records are representations of ideas or of objects, he wrote, but objects themselves can be regarded as "documents" if you are informed by examination of them. He cites, as examples, natural objects, artifacts, objects bearing traces of human activity (such as archaeological finds), explanatory models, educational games, and works of art (p. 217). Elsewhere he wrote of microscopic slides and museum objects as essentially documentary in character (Otlet 1990, 153 and 197). It should be clear that *document* is being used here in a broad, inclusive way, subsuming more precise terms, such as *manuscript* or *specimen*.

Defining *document* is the starting point of Suzanne Briet's long-neglected manifesto of 1951, *Qu'est-ce que la documentation?* [What is documentation?]. A document, she wrote, is evidence in support of a fact and could be any physical or symbolic sign, preserved or recorded, intended to represent, to reconstruct, or to demonstrate a physical or conceptual phenomenon. Documentation should not be viewed as being concerned with texts, she declared, but with access to evidence (Briet, 1951, 7; 2006, 9-10). In a famous example she declared that an antelope in the wild was not a document but if captured, placed in a taxonomy, and exhibited into a cage it has been made into a document (Frohmann 2011). These ideas were presented and more carefully explained by Robert Pagès, a student in Briet's educational program for documentalists, three years earlier in a paper that seems to have been overlooked (Pagès 1948).

Many writers include maps, diagrams, drawings, and images with textual records in a broader category of *graphic records*. The inclusion of non-graphic items (plant specimens, animals, and other objects) has found less acceptance. Pagès argues that a textual or graphic document is always about, or refers to, some other concept or entity and, in that sense, is always secondary to it. But a non-graphic object (e.g. a gorilla in a cage, a mineral specimen, Napoleon's hat) is not about anything else, it is simply itself, an "autodocument". It becomes meaningful only in conjunction with other symbols and (secondary) documents and in this way it bridges the gap between bookish learning and lived experience (Pagès 1948; see also Frohmann 2011).

Two parallel developments to this modern expansion of the notion of a document can be noted. One was the rise of historical-critical textual studies (initially called "higher criticism") in philology, especially in Biblical studies, in which contextual, cultural, and other influences are considered in contrast to textual criticism ("lower criticism") focused on a text itself and variant versions (witnesses). The other was a transformation of the writing of history when the rigorous "scientific" approach focused in the authenticity of archival records was displaced by the work of

historians associated with the *Annales* journal who accepted a wider range of documentary evidence: anything that can be interrogated by a historian is acceptable as evidence.

The broad view of *document* arose again in the 1990s. For example, in Norway the law requiring legal deposit of a copy of all published documents in the national library had been extended beyond printed material to include all media, which would require a broader range of expertise among librarians and led to the founding of a new educational program led by Niels W. Lund at the University of Tromsø in 1996. The program took the notion of a document as central, examined its physical, social and cognitive aspects, and adopted the name *documentation studies* (*dokumentasjonsvitenskap*) (Lund 2007). Already in 1991, Buckland, seeking to accommodate natural history museum specimens within an inclusive view of information science, had sorted uses of the word *information* into three categories: as knowledge imparted, as process, and as thing. The last ("information as thing") corresponds with a broad view of *document* (Buckland 1991; 1997). "Document" is thus used as a generic term which includes not merely published and unpublished books, articles, and letters, but also music, pictures, and sound recordings. *Document* is not limited to texts, but can include, at least in theory, museum objects, animals in a zoo, even a landscape (Grenersen, Kemi and Nilsen 2016).

Related issues were taken up in a wide-ranging examination led by Roger T. Pédauque (a collective pseudonym) of what could be learned from the shift from print media to digital media (Pédauque 2003; 2006; 2007). Pédauque distinguished three aspects of document use: physical perception (seeing), intellectual effort (reading), and interpretation (understanding), in French: *vu*, *lu*, *su*. Recently, a concise book-length introduction to document theory is provided by *Le document: communication et mémoire* [The document: Communication and memory] by Tricot, Sahut and Lemarié (2016).

The primary institutional context for Otlet and Briet was the International Institute for Bibliography (IIB, later named the International Federation for Documentation, FID), and, for Briet, the Union Française des Organismes de Documentation (UFOD). Lund and others created the Document Academy, an informal collective, which, since 2001, has organized numerous conferences, workshops, and, now, published proceedings (Lund and Buckland 2008; Buckland and Lund 2013; Proceedings 2014; Skare, Lund and Vårheim 2007). In parallel, within of the Association for Information Science and Technology the Special Interest Group in the History and Foundations of Information Science has hosted numerous discussions since the early 1990s.

Of course many others, including art historians, archivists, engineers, lawyers, media specialists, historians, and textual scholars, also have specialized interests in aspects of documents.

3. Status as a document

An object is considered to be a document when there is an assertion or a perception of evidence for some belief. The effect of a document, then, depends on belief concerning some aspect of reality. Of course, any assertion or belief may be deemed by others (or by the same person at another time) to be erroneous, out-of-date, incomplete, and/or a misrepresentation. However, to function as a document requires an act of perception (reading, viewing, or otherwise sensing), so a perceiver is as necessary as a creator.

Meyriat distinguished two kinds of document: a document by intention (i.e. created to be a document) and a document by attribution (i.e. regarded as a document) (Tricot, Sahut and Lemarié 2016, 16). One can further distinguish two sorts of attribution: by a creator and by whoever perceives (reads, views, senses) the document (Buckland 2014).

Three origins of document can be identified:

(i) A document can be *created as* a document: written, drawn, or otherwise made as a document, ordinarily producing an inscription on a flat surface. This is a conventional view.

(ii) Objects can be *made into* or presented as a document. This is a functional view.

(iii) Any object, whether or not included in i or ii, may be *regarded as* a document by an perceiver, whether or not its creator, if any, intended it to be a document. This is a semiotic view.

4. Document relationships

Documents are commonly compared with, or contrasted with, other concepts. There is a long tradition of differentiation, dichotomizing, for example, document from data, documentation from bibliography, or librarianship from information science. Ørom (2007), for example, compares and contrasts uses of document and information. Although often helpful as a starting point, this approach, by emphasizing differences, has the disadvantage that it tends to simplify, to impose restrictive views on one or both of the pair, and to understate what is in common.

4.1. Document and data

Documents are often contrasted with data. Commonly this is simply a matter of format. An extended textual record is considered a document and short numeric records are referred to as data. Similarly, *document retrieval* has sometimes been used to denote the selection of entire articles, books, or reports from a collection and *data retrieval* to refer to the selection of fragments from within one or more documents. This is a convenient part-whole distinction.

Furner (2016) provides a thorough historical and analytical examination of distinctions between data, document, and information. He concludes that most usages are unsatisfactory and that, "It is not in fact the case that documents are made up of data, nor that the document is a species of dataset: rather it is the other way around, in both respects. A dataset is made up of documents; and the dataset is a species of document" (p. 303).

4.2. Documents and facts

The evidential role of a document leads naturally to discussion of facts. There is a contrast between Paul Otlet and Ludwik Fleck. Otlet saw documents as statements of facts representing details of the world such that an assemblage of documents could constitute a mirror of the world (Frohmann 2008). At the same time Ludwik Fleck argued that concise statements of fact were misleading because they are necessarily simplified. Writings needed to be understood within the

cognitive and cultural context of the author which may be quite different from that of the reader (Fleck 1935/1979; Cohen and Schnelle 1986).

4.3. Documents and texts

Not all documents are textual, but most commonly documents are of interest because of text inscribed in or on them. It is easy to treat text and document as interchangeable, but the distinction is important. The study of texts (philology) makes a distinction between the study of the text itself in isolation (which used to be called "lower criticism") and study of a text in its material, social, and historical context ("higher criticism"). While there can be no objection to the study of a text in isolation, document theory is very much concerned with the material, historical, and cultural contexts of texts in the traditions of the textual scholar Jerome McGann (1983) and the bibliographer Donald F. McKenzie (1986). But, as McKenzie points out, a document can be a signifying non-abstract object without any form of writing.

4.4. Documents and works

In the context of knowledge organization, the term *work* has been used with two quite different meanings. In one meaning, *work* refers to a material object, typically a printed book ("the work in hand"), in which case a work is a document. But *work* is also used in an abstract sense to denote an intellectual product. In this second sense a *work* and a *document* cannot be the same. They are different in kind, one immaterial and the other material. They are closely related, however, because a work can be expressed physically (and so made accessible) only in the form of a document.

In the terminology of the Functional Requirements for Bibliographic Records (FRBR) a document (item) is an instance (token) of a manifestation (type) and one need not depend on the conjectured notions of expression and work (IFLA 1998).

4.5. Documents and information

As noted above, the word *information* is used with a number of different meanings. One important usage is to refer to physical stuff: Data, records, inscriptions, and so on. This usage ("information-as-thing") corresponds to ordinary notions of document. Other usages do not. Ørom (2007) argues for treating document and information as simply different. Frohmann's Deflating information: from science studies to documentation (2004) addresses the relationship between information and documents thoroughly.

4.6. Documents and knowledge

The word *knowledge* is also used with multiple meanings. As with *work*, we can distinguish both abstract and material uses. What we *know* (i.e., believe with some confidence) is in our minds and should not be considered a document. But in practice the term *knowledge* is often also used to denote recorded knowledge, typically in material form as texts, diagrams, or other graphic form. In this second, extended sense, *recorded knowledge* refers to documents.

4.7. Documents and signs

An object is considered a *document* if it does, or could, reveal or signify something. A document is expected to be actually or potentially meaningful. As such it is a kind of sign, as Briet stated explicitly when she defined *document* as "any concrete or symbolic indexical sign, preserved or recorded toward the ends of representing, of reconstituting, or of proving a physical or intellectual phenomenon" [Tout indice concret ou symbolique, conservé ou enregistré, aux fins de représenter, de reconstituer ou de prouver un phénomène ou physique ou intellectuel] (Briet 1951, 7; 2006, 10).

4.8. Relationships between documents

The meaning of a document—what it is perceived as signifying—makes a document important, but what it signifies is influenced and extended by relationships with other documents and with people. Any interaction or any shared attribute can be used to indicate a relationship, so the scope for establishing relationships is, in effect, unlimited, including textual relationships between different witnesses of texts, literary relationships between different treatments of the same theme, semantic relations (e.g. a movie and its poster; a book and its reviews), bibliographical and citation links, topical similarities, and so on. Consequently, set theory, graph theory, and bibliometrics can find wide application in document relationships.
There are also part-whole relationships. For example, an article may include an illustration which may also serve separately as a document in its own right as well as being a part of the larger article. Lund uses *doceme* for such as part (Lund 2004, 99).

5. What documents do

Documents are ordinarily seen as mitigating the effects of time on memory and of space on communication. Here we examine roles of documents.

5.1. Documents as communication

Documents are transmitted across space and between people. The rather idealized approach of Tricot, Sahut and Lemarié (2016) is based on the conversation theory of Paul Grice as developed by Wilson and Sperber (2012). The rapid increase in telecommunication technologies greatly facilitates and increases the communication of documents.

5.2. Documents as communication across time

Otlet and many others have seen documents and documentary systems as constituting a kind of aid for or extension of human memory. Tricot, Sahut and Lemarié (2016) enumerate problems of memory and the ways in which documents may aid memory. Problems of human memory include:

1. Forgetting: A document might lead to recovering memory or learning again.

2. Distraction: Not paying attention and so not remembering when needed, might be remedied by a "to do" list.

3. Long and medium-term blocking: It is easier to recognize something on a list than recall it without a prompt.

4. Misattribution, suggestibility, and bias: Confused or false memories may arise because we try to make sense when we recall. Reference to a document aids consistency.

5. Persistence: Some memories haunt us. You can destroy a document but not a memory.

Documents, then, not only allow the preservation and transmission of records over time, they serve memory by preserving, verifying, allowing recognition of what was forgotten, and anticipating future needs.

5.3 Documents as cognitive resources for learning

The unifying treatment of document theory in relation to communication and memory by Tricot, Sahut and Lemarié (2016) does not exhaust the possibilities. For example, when I examine a document for the first time I am not remembering it nor simply receiving a transmitted communication, but, rather, developing my prior understanding and generating new ideas. Although the creator of the document probably had a communicative intent, that intent may be not be clear to me, or not persuasive, or not of interest. I am engaged in some purposive cognitive effort. My intent and the outcome could be regarded as learning because I am becoming familiar with what others already knew and discovering what was apparently not already known. A scholar uses evidence perceived in documents as an ingredient to develop new ideas.
There is, therefore, a trinity of transmission (communicating), recording (documenting), and learning (becoming informed). Lund has stressed that these three activities constituting physical (material), cognitive, and social (cultural) dimensions are always more or less co-present, a feature he terms *complementarity* (Lund 2004, 96-97; Skare 2009; Olsen et al. 2012).

5.4. Agency

To ask what documents do or to speak of documents having "agency" is to risk figurative language, but documents act as agents in the sense that their existence and features do have material consequences in enabling (affording) outcomes and further actions. The notion of agency, popularized in actor-network theory, is problematic since in ordinary English *agent* implies a conscious individual acting responsibly. In actor-network theory (as in *chemical agent* or *cleaning agent*) an agent is anything whose existence and status has consequences, which could include gravity, an acid, a hammer, or a speed bump in the road (cf., Latour 2007). With this broad definition one can consider documents as agents.

Consider a single, simple document: a library catalog record. It is itself a document, but it also represents some other larger document, usually a book. Its representing can be seen as having three forms: It is representational in the sense that it describes the book; it is indexical in the

sense that it points to the shelf location of the book; and it may serve as a surrogate for the book if, for example, one merely needs to verify the date of publication or the spelling of the author's name and one is willing to trust the reliability of the catalog record. And, by extension, the set of catalog records (the catalog) represents the collection.

From another perspective a catalog represents the librarian's knowledge. An individual record represents (some of) what the librarian knows about that book, but it is likely to be an incomplete representation because the librarian may know, but the record does not show, that this book has been discredited or superseded by some other book, which may not be in the library, or that the publisher is a vanity press. Nevertheless, the catalog record is representing and acting for the librarian. Again, by extension, the catalog as a whole reflects what the librarian knows about the collection as a whole (Buckland 2017b).

5.5. Instruments

The issue of documentary agency arises in the philosophy of science and technology. Davis Baird identifies three types of material objects that embody "knowledge": (i) Models, such as a model airplane, which show a resemblance; (ii) Devices that can create a phenomenon, such as a dynamo; and (iii) Measuring instruments which bridge the gap between something (*noumenon*) and what we can perceive (*phenomenon*) (Baird 2004; Bunge 2010).

A favorite example is an orrery, an eighteenth century mechanical device which models the movement of the moon around the Earth and the movement of the Earth and other planets around the Sun. As a result, an orrery can be used to show the movement of, say, the Earth's moon relative to the Sun or to Mars, which would have been difficult by any other method before the development of modern digital computation. This corresponds to Otlet's inclusion of educational toys within his definition of documents. Mechanical models are now largely replaced by digital simulations, but that is a difference in technology not of function.

Baird argues that material devices of this type are tools that facilitate human thought in a manner that is functionally comparable to theories. The former are material and the latter abstract but they have comparable roles.

6. Document, evidence, and experience

Otlet, Briet, and others have focused on the role of documents as providing factual, truthful evidence, but reality is far more complex. Not all knowledge or teaching is factual and not all truth claims are valid. Fictional narratives are used to teach moral lessons. The wider senses of *docere* (show, tell, demonstrate) are rhetorical and like rhetoric may not reduce to facts or factual evidence. An emphasis on evidence does not satisfactorily account for all uses of books, musical performances, or other kinds of documents.

Consider a volume of Aesop's fables, fanciful stories of the activities of animals that behave like humans, usually foolishly, followed by a statement of moral advice. The moral advice qualifies the fable as teaching within the spirit of *docere*. In medieval Europe, a four level classification was used for didactic texts. The basic level was the literal sense of a text, the ordinary level of

factual documents, such as a software manual. The symbolic activities of the animals in Aesop's fables are interpreted at a second, figurative, allegorical level. A third, tropological level, in which the reader infers where her or his moral duty lay, is made explicit in the advice (e.g. Be careful what you wish for) following the story in an Aesop's fable. And since religious thought can be apocalyptic and visionary, suitable texts inspire meditation at an anagogic level: Reading moralizing fables and parables should make one wiser and more compassionate (Wikipedia 2017a). All of these four levels are consistent with the teaching role implied by the root of *docere*.

A rather different situation arises when little or no lesson appears as with light fiction, amusing drawings, or frivolous operas. These cases go beyond ordinary understandings of *evidence* even though something is revealed. Books are commonly enjoyed for amusement or as escapist reading and one may choose to reject light fiction and comparable entertainment in other media as "documents" or simply not call them documents. There are two options. One choice is to decide that what may be a document in form (a book or film) but when lacking credible claims to truth, legal validity, or moral correctness should not be considered a document. A second choice is to retreat from insistence on a strong sense of evidence (factual proof) and accept that narrative telling and emotional affect, even without claims to truth or moral correctness, are on a continuum with the examples already considered and so not incompatible with the *telling* implied by *docere*. The situation is not simple because, like Aesop's fables, any frivolous story is liable to have some figurative, aesthetic, cathartic, or relaxing effect.

Frohmann (2011) suggests that Briet's interest in documents as evidence derives from her concern with scholarship and he argues persuasively that an emphasis on the evidential role of documents misses the real significance of her analysis.

> The deeper point about Briet's antelope is that something becomes a document by virtue of its arrangements with other things, not about a privileged form of those arrangements, such as their evidentiary functions. I take from Briet the idea that in complex arrangements *things* exercise documentary agency, which is capable of being detected, understood, and engaged in many different ways, and by many different kinds of actors, both human, and nonhuman. The problem is to show how a thing's documentary agency, power, or force – which I call its documentarity – is exercised by virtue of those arrangements. [...] [H]ow do writing, traces, and documentation emerge from the interactions between a *thing* – this antelope – and other elements of its specific arrangements? (pp. 173-174).

In other words, a document may make something *evident*, i.e. show something. This is a weaker, but much broader statement than is implied by *evidence*, which indicates a narrower, more forceful impact in, say, a scientific conclusion or a legal process.

The idea that a document can reveal (or has documentary force for) some perception has a corollary: Looking outwards from document to beliefs about reality brings one into contact with the many other fields engaged in examining reality. This is consistent with views that documentation (or information science) is in some sense a meta-discipline.

7. Physical and cognitive and social aspects

By now it will be clear that discussions of documents include physical (material) aspects, cognitive, and social (cultural) aspects (Lund 2004, 96-97; Buckland 2016).

7.1 Physical aspects

There is necessarily a physical aspect to documents. ("Physical" seems preferable to "material" because it is more hospitable to the inclusion of movement, gesture, and performance.) The greatly varying physical characteristics of media (e.g. clay tablets, papyrus, paper, microfilm, and the many forms of digital documents) and the varied and powerful techniques used (notably language, writing, printing, copying, and, especially, digital manipulation) lead to great variety in the form, affordances, stability, and longevity of documents. These interesting and important matters are, however, beyond our present scope.

7.2. ...and cognitive

There is also necessarily a cognitive aspect: If anything appeared to lack actual or potential meaning, we would not regard it as a document.

The standard view is a heroic tale of how documents and ever-improving technology progressively mitigate the constraints of time and space on communication. This view is not wrong, but it is incomplete, because movement through time and across space constitutes a change in context, in cultural context as well as spatio-temporal context. No document is ever perceived outside of any context and the context affects perception and interpretation. The attention paid to searchers in relation to their contexts needs to be accompanied by attention to documents in relation to their contexts.

That documents have cognitive as well as physical aspects has an important methodological consequence. Only the physical aspect of documents can be treated scientifically in the normative sense of the physical sciences, which do not extend to perceptions of meaning. This is why *relevance* as an evaluation criterion in information retrieval evaluation is so elusive. We know what it is but it resists satisfactory definition or quantification (White 2010). This situation is in sharp contrast with Shannon's Information Theory which can be treated scientifically precisely because it excludes meaning and has no cognitive aspect. It is the presence of a cognitive aspect of documents that ensures that any understanding of information science that includes learning and understanding must have one foot in the humanities and qualitative social sciences.

7.3. ...and social

Documents play social roles (Brown and Duguid 2017) and have multiple social aspects. As McKenzie (1986) has reminded us, the physical production of documents requires multiple social actors. Publication of a book requires, in addition to an author, a publisher, a printer, manufacturers of paper and ink, binder, booksellers, an elaborate infrastructure of transportation, financial systems, and, of course, readers. Documents are also subject to a wide variety of

socially imposed regulations relating, for example, to privacy, security, sedition, blasphemy, standards, and intellectual property rights.

Perceiving meaning and becoming informed is essentially a cognitive development by a living individual, yet society depends on the communication of meaning *between* individuals. It is here in the shared *intersubjective* understanding that documents play a crucial role. One person learning from another requires that the subjective understanding of one be physically revealed somehow, through a gesture, perhaps, or speech, or writing – through a document – so that one or more others can then perceive it and try to make sense of it. This process, the *social* construction of meaning, enables communication, and so collaboration, and thereby culture and society. (For a convenient introduction, see Zerubavel 1997; also Mannheim 1936, chap. 1 and Berger and Luckmann 1966).

In addition, the meanings of documents derive from shared cultural codes, language in a broad sense. As Fleck (1935/1979) emphasized, to be understood, a text needs to be understood in terms of the cultural context and mindset of the writer and yet it will be read within the cultural context and mindset of the reader. Difficulties arise as cultural distances between these contexts and mindsets increase.

7.4. Reversing the view

Document theory has a document-centered perspective and looks outwards to see how documents are engaged in physical, social, and cognitive worlds. So, if we are concerned with the relationships between documents and reality, we should consider reversing our views of these relationships. What if we were to look inwards from those three environments and observe how documents figure in our perspectives?

In terms of the physical environment, we would be concerned with how technology and material resources are, or might be, used in documentary activities. As one important example, the dramatic rise of data technologies is reversing the relationship between data sets and individuals: Instead of data as a useful resources for human self-actualization, data sets are increasingly establishing human identities for each of us (Day 2014a;b). Note that document theory differs from theorizing information in that use of the word *information* has been extended by some theorists to include physical forms, forming, and patterns that have no direct relationship with cognitive activity, with humans becoming informed. One sees this in Shannon's Information Theory and the widespread use of the phrase "Information Technology" to include uses of electronics unrelated to human knowing.

If we start with a view from social studies and social history, we may hope to acquire insight into the roles that documents play in our social lives, the relative importance of different document-related practices in diverse circumstances, what changes have occurred over time, what the future might bring, and how (and why) we might seek to influence future documentary developments.

From a cognitive view, from a concern with learning and mental work, what roles do (or might) documents play in acquiring the knowledge we seek? Learning how to make good use of

documents is major component of learning how to learn and closely related to the concerns of knowledge organization. Vladimir Stibic, a highly-educated documentation specialist, wrote that he had learned many interesting and useful facts in his formal education but:

> I do blame all my schools because they never tried to teach me how to learn and how to work. [...] [S]tudents, though well prepared as concerns general knowledge and their specialist fields, are given no advice on how to work efficiently, how to organize their work, how to read, listen and study, how to gather, store and organize information, or how to use modern technical means to save time and effort and to improve their own productivity (Stibic 1982, vii).

The solution, Stibic explains, is to teach the tools and know-how useful for efficient and effective learning and mental work and to a significant extent that implies understanding documents and documentation.

Good practice in librarianship includes preparing pathfinders, very concise topical guides to documentary sources that are locally available. There is also the genre of "How to find out" manuals that usually take a broader view of a single subject area. But this is only one aspect of learning to learn and it needs to be complemented by attention to critical thinking, the scientific method, elementary statistics, the making and organizing of notes, how to cite, technical writing, the use of technical tools (such as spreadsheets and word-processing), and so on. This range includes but goes beyond "information literacy". Didactic materials are available on each of these and digital computing has opened additional opportunities for "tools of the mind" in the tradition of Douglas Engelbart and others (Wikipedia 2017b).

These components exist as resources generally isolated from each other. Document theory suggests a basis for a coherent approach to the tools and know-how for learning and discovery. The nearest to such a unified approach appears to be the "hodegetic" (German: *Hodegetik, Wegweisung*, showing the path) tradition in central Europe of the seventeenth to early twentieth centuries in which scholars, especially scholar librarians, prepared more or less inclusive guides with titles like *Die Technik der geistige Arbeit* [Tools and know-how for cognitive work]. Schmidmaier (1970) provides an annotated bibliography and introduction. A late example of this genre in English is Stibic's *Tools of the mind: Techniques and methods for intellectual work* (1982), which has an emphasis on office equipment. It was inspired by the lectures on how to work, emphasizing mental work, by the Czech leader T.G. Masaryk (1898/1938).

More narrowly, document theory offers a conceptual framework for curriculum, faculty selection, and professional education in knowledge organization, librarianship, information science/studies, and related areas.

8. Methods

In his "Revisiting 'What is a document?'" Bernd Frohmann (2009, 296) argued for narrative qualitative analysis. He recommended moving away from a preoccupation with definitions and beginning instead with clear cases of things we agree to be documents or activities we agree to be instances of documentation, and then telling stories in which new cases are introduced by

analogy, similarity, and resemblance. He illustrates this approach by examining the documentary performance of the cabinets of curiosities that were fashionable in sixteenth-century Europe. The strange objects collected were, both individually and collectively, well outside the conventional view of what a document is, yet they were used to signify social meaning.

In addition to considering possible documentary properties of objects not ordinarily considered documents, we also need to consider how our ideas about documents may need to evolve to accommodate electronic technologies. For example, a passport, an archetypal document, is used to allow one to cross a barrier and enter new territory. Examining that role leads smoothly to the role of an electronic card keys (plausibly also a document of sorts) that allows one to enter a locked building and on to a metal key that arguable has the same documentary function but is outside any ordinary sense of "document" (Buckland 2014). Lund has made extensive use of qualitative analyses (Lund in preparation).

We may hope that the qualitative approaches used to develop document theory will eventually be joined with more empirical, quantitative analysis. Given the pervasive role of documents in society the benefits of quantitative analyses are likely to be impressive. For example, Hernando De Soto (2000) estimated that correcting weaknesses in the documentation of real property ownership in developing and formerly communist countries could increase the economic value of these properties by some US$ 8 trillion (sic). Of course, achieving that improvement in the documentation of ownership would depend on there being an adequate legal regime, a dependable judicial system, law and order, efficient government administration, and political will. As with passports, documents are effective only within a competently organized social context, but the potential for generating wealth by documentary means is enormous. Similarly, the negative consequences of the economic recession of 2008 were largely generated by the widespread use of dishonest mortgage documents and derivatives of them (Dayen 2016). On an even larger scale Max Boisot explains political structures, notably the prevalence of inefficient crony capitalism (as distinguished from market capitalism) in terms of the degree of trust that recorded statements will be honored (Boisot 1998; Boisot 2011, chap 2; Wang and Buckland 2016). More familiar is the widespread misuse of links between documents in bibliometric and scientometric analyses as a substitute for honest evaluation (Gingras 2016).

9. Relationships with other approaches

Document theory is associated with the practice of documentation and, therefore, with bibliography, information science, and librarianship. In continental Europe, self-described documentation centers were commonly part of a library's service. Following a tour of the USA in 1951-1952, Briet, herself a professional librarian who advocated documentation, addressed the relationship between librarianship and documentation. She considered that librarians and documentalists were not different in kind, but in emphasis. Librarians manage collections and develop bibliographical apparatus; documentalists focus in advancing the intellectual work of the population they serve. Differences in type of material selected, forms of indexing, and timeliness flow from that difference in emphasis. She noted that in the USA the term *documentation* was not much used but that the practice was flourishing as special library services (Briet 1954; Buckland 1996). Similarly, Mortimer Taube (1952) observed that documentation, by including attention to the creation and publication of documents, had a broader scope than librarianship. In

English-speaking countries use of *documentation* was replaced by *information science* (or *management* or *studies*), but these changes of name do not diminish the relevance of document theory.

More broadly, if we take documents as a starting point and look outwards into how they are engaged in the physical, social, and cognitive aspects of our world, affinities and overlaps are revealed not only with media studies, but also many other areas including anthropology (e.g. Riles 2006), performance studies (e.g. Taylor 2003), philosophy (e.g. Ferraris and Caffo 2014), publishing (e.g. Bhaskar 2013), and across the humanities (e.g. Gitelman 2014, 189-204) and beyond.

10. Commentary

We have followed previous writers, notably Otlet, Briet, Lund, and Frohmann, in exploring how documents are involved in our minds, in society, and in the physical world. To do this we have sought to avoid or defer limits based on physical format or definitions that differentiate *document* from, say, *data* or *information*. This approach reveals a very rich and complex landscape in which documents and documentary roles are deeply engaged in many aspects of society, technology, and knowledge.

Documents are an integral part of how we learn, what we know, the division of labor, our sense of identity, and our wealth. It is a landscape in which documents and documentary roles overlap with many other disciplines and is characterized by continua rather than by separations. Fences are built by disciplines, not by explorers and should be regarded with some skepticism. In this view document theory overlaps and connects with many other fields: education, communication, economics, media studies, politics, telecommunications, and more.

If we accept this "rich landscape" perspective then we can draw some conclusions:

1. Some areas of documentation are so closely overlapping with other fields as to make separation artificial. For example, with socially-aware approaches to bibliography and historical-critical textual studies.

2. Awareness of continua in the landscape suggests new opportunities and applications. As one example, document theory appears to provide a promising conceptual framework for the important pragmatic field of information and communication technologies (ICT). The ICT revolution needs to be seen as a document revolution enabling one to relate to others or oneself by reducing constraints of space and time (Tricot, Sahut and Lemarié 2016, 133).

3. The document-centric perspective of document theory appears to hold promise for a unifying and coherent approach to the diverse tools and methods available as aids to learning.

4. Document theory provides a basis for reviewing completeness of the curriculum for schools of information and, more broadly, of the technologies and methods for learning.

5. Documentary regimes reflect social evolution (Buckland 2017a; Tricot, Sahut and Lemarié 2016, 135-136).

6. What one chooses to call a "document" is a pragmatic decision and should not be allowed to obscure the underlying relationships discussed above. Any working definition of *document* may be useful within certain situations, but remains an arbitrary matter of convenience, a "word game".

7. Significant intellectual developments in information studies are rare. The working-out of the implications of moving from a format-based view (documents as textual or graphic objects of specific kinds) to the phenomenological approach of Pagès and Briet, is one of the most important conceptual developments at this time.

Resources

Convenient introductions to document theory have been provided by Lund (2009, with a shorter version by Lund and Skare 2010), by Tricot, Sahut and Lemarié (2016), and Buckland (2015). A wider range of writings on document theory, mostly generated under the auspices of the Document Academy, can be found in Skare, Lund and Vårheim (2007) and the Proceedings from the Document Academy (2014-).

References

Baird, Davis. 2004. *Thing Knowledge: A Philosophy of Scientific Instruments*. Berkeley: University of California Press.

Berger, Peter L. and Thomas Luckmann. 1966. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Garden City, NY: Doubleday.

Bhaskar, Michael. 2013. *The Content Machine: Towards a Theory of Publishing from the Printing Press to the Digital Network*. New York: Anthem Press.

Boisot, Max H. 1998. *Knowledge Assets: Securing Competitive Advantage in the Information Economy*. Oxford: Oxford University Press.

Boisot, Max H. (Ed.) 2011. *Collisions and Collaboration: The Organization of Learning in the ATLAS Experiment at the LHC*. Oxford: Oxford University Press.

Briet, Suzanne. 1951. *Qu'est-ce que la Documentation?* Paris, EDIT.

Briet, Suzanne. 1954. "Bibliothécaires et Documentalistes". *Revue de la Documentation* 21, no. 2: 41-45.

Briet, Suzanne. 2006. *What is Documentation?* English translation of the classic French text. Lanham, MD: Scarecrow Press. Also http://info.slis.indiana.edu/~roday/what%20is%20documentation.pdf

Brown, John Seely and Paul Duguid. 2017. *The Social Life of Information*. Updated, with a New Preface. Boston: Harvard Business Review Press.

Buckland, Michael K. 1991. "Information as Thing". *Journal of the American Society of Information Science* 42:5: 351-360. http://people.ischool.berkeley.edu/~buckland/thing.html

Buckland, Michael K. 1996. "Documentation, Information Science, and Library Science in the U.S.A". *Information Processing & Management* 32, no. 1: 63-76.

Buckland, Michael K. 1997. "What is a "Document'?" *Journal of the American Society for Information Science* 48, no. 9: 804-809.

Buckland, Michael K. 2014. "Documentality Beyond Documents". *The Monist* 97, no 2: 179-186. http://escholarship.org/uc/item/55v7p74x

Buckland, Michael K. 2015. "Document Theory: An Introduction". In Mirna Willer, Anne J. Gilliland and Marijana Tomic (Eds), *Records, Archives and Memory: Selected Papers from the Conference and School on Records, Archives and Memory Studies, University of Zadar, Croatia, May 2013* (pp. 223-237). Zadar: University of Zadar, 2015. ISBN 978-953-331-080-0. Preprint at http://people.ischool.berkeley.edu/~buckland/zadardoctheory.pdf

Buckland, Michael K. 2016. "The Physical, Mental and Social Dimensions of Documents". Proceedings of the Document Academy 13, issue 1, article 4. http://ideaexchange.uakron.edu/docam/vol3/iss1/4/

Buckland, Michael K. 2017a. *Information and Society*. Cambridge, MA: MIT Press.

Buckland, Michael K. 2017b. "Library Technology in the Next Twenty Years". *Library Hi Tech* 35, no 1: 5-10. Also http://escholarship.org/uc/item/9gs9p655

Buckland, Michael K. and Niels Windfeld. 2013. "Boyd Rayward, Documentation, and Information Science". *Library Trends* 62, no. 2: 302-310. http://muse.jhu.edu/article/540515/pdf http://people.ischool.berkeley.edu/~buckland/bucklandlund.pdf

Bunge, Mario. 2010. "Reading Measuring Instruments". Spontaneous Generations: A Journal for the History and Philosophy of Science 4: 85-93. http://spontaneousgenerations.library.utoronto.ca/index.php/SpontaneousGenerations/article/view/11725/11222

Cohen, Robert S. and Thomas Schnelle (Eds). 1986. *Cognition and Fact: Materials on Ludwik Fleck*. Dordrecht: Reidel Publishing Co. (Boston Studies in the Philosophy of Science 87).

Day, Ronald E. 2014a. "The Data—it is Me!" ("Les données—c'est Moi'!). In: *Beyond Bibliometrics: Metrics-based evaluation of research*, ed. by Blaise Cronin and Cassidy R. Sugimoto. Cambridge, MA: MIT Press, 67-84.

Day, Ronald E. 2014b. *Indexing It All: The Subject in the Age of Documentation, Information, and Data*. Cambridge, MA: MIT Press.

Dayen, David. 2016. *Chain of Title: How Three Ordinary Americans Uncovered Wall Street's Great Foreclosure Fraud*. New York: New Press.

Ferraris, Maurizio. 2013. *Documentality: Why It Is Necessary to Leave Traces*. New York: Fordham University Press.

Ferraris, Maurizio and Leonardo Caffo (Eds.) 2014. "Documentality". *The Monist* [Special issue] 97, no 2: 159-255.

Fleck, Ludwik. 1935/1979. *Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv*. Basel: Schwabe. Transl. as *Genesis and development of a scientific fact*. Chicago: Univ. of Chicago Press, 1979.

Frohmann, Bernd P. 2004. *Deflating Information: From Science Studies to Documentation*. Toronto: University of Toronto Press.

Frohmann, Bernd P. 2008. "The Role of Facts in Paul Otlet's Modernist Project of Documentation". In *European Modernism and the Information Society*, ed. by W. Boyd Rayward. Aldershot, England: Ashgate, 75-88.

Frohmann, Bernd P. 2009. "Revisiting "What is a document?"" *Journal of Documentation* 65, no. 2: 291-303.

Frohmann, Bernd P. 2011. "The Documentality of Mme Briet's Antelope". In: *Communication Matters: Materialist Approaches to Media, Mobility and Networks*, ed. Jeremy Packer and Stephen B. Crofts. Wiley. London: Routledge, 173-182.

Furner, Jonathan. 2016. "'Data': The Data". In: *Information Cultures in the Digital Age: a Festschrift in honor of Rafael Capurro*, eds M.Kelly, J. Bielby. Wiesbaden Springer, 287-306.

Gingras, Yves. 2016. *Bibliometrics and Research Evaluation: Uses and Abuses*. Cambridge, MA: MIT Press.

Gitelman, Lisa. 2014. *Paper Knowledge: Toward a Media History of Documents*. Duke University Press.

Grenersen, Geir, Kjell Kermi, and Steinar Nilsen. 2016. "Landscapes as Documents: The Relationship between Traditional Sámi Terminology and the Concepts of Document and Documentation". *Journal of Documentation* 72, no. 6: 1181-1196.

IFLA Study Group on the Functional Requirements for Bibliographic Records (1998). *Functional Requirements for Bibliographic Records*. München: K.G. Saur.

Latour, Bruno. 2007. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford, UK: Oxford University Press.

Lund, Niels Windfeld. 2004. "Documentation in a Complementary Perspective". In: *Aware and Responsible: Papers of the Nordic-International Colloquium on Social and Cultural Awareness and Responsibility in Library, Information, and Documentation Studies (SCARLID)*, ed. by W. Boyd Rayward. Lanham, Md. Scarecrow Press, 93-102.

Lund, Niels Windfeld. 2007. "Building a Discipline, Creating a Profession: An Essay on the Childhood of 'Dokvit.'" In: Roswitha Skare, Niels Windfeld Lund, and Andreas Vårheim, eds. *A Document (Re)turn: Contributions from a Research Field in Transition*. Frankfurt a. M.: Peter Lang, 11-26.

Lund, Niels Windfeld. 2009. "Document Theory". *Annual Review of Information Science and Technology*, 43, 399-432.

Lund, Niels Windfeld. 2010. "Document, Text and Medium: Concepts, Theories and Disciplines". *Journal of Documentation* 66, no. 5: 734-749.

Lund, Niels Windfeld (in preparation). *Introduction to Documentation Studies*. London: Facet. (Foundations of the Information Sciences). Announced to appear December 2018.

Lund, Niels Windfeld and Michael Buckland. 2008. "Document, Documentation, and the Document Academy". *Archival Science* 8: 161-164.

Lund, Niels Windfeld and Roswitha Skare. 2010. "Document Theory". In *Encyclopedia of Library and Information Sciences*. 3 rd ed. Marcia J. Bates and Mary Niles Maack (Eds.), Boca Raton, FL: Taylor & Francis, Vol II: 1632-1639

Mannheim, Karl. 1936. *Ideology and Utopia: An Introduction to the Sociology of Knowledge*. New York, Harcourt, Brace & World.

Masaryk, Thomas Garrigue. 1898/1938. *Jak pracovat?* Prague: Cin, 1898. English translation in: *The Ideals of Humanity; and, How to Work: Lectures Delivered in 1898 at the University of Prague*. London: Allen & Unwin, 1938.

McGann, Jerome J. 1983. *A Critique of Modern Textual Criticism*. Chicago University of Chicago Press.

McKenzie, Donald Francis. 1986. *Bibliography and the Sociology of Texts*. London: British Library.

Olsen, Bernt Ivar, Niels Windfeld Lund, Gunnar Ellingsen, and Gunnar Hartvigsen. (2012). "Document Theory for the Design of Socio-Technical Systems: A Document Model as Ontology of Human Expression". *Journal of Documentation* 68, no. 1: 100-126.

Ørom, Anders. 2007. "The Concept of Information versus the Concept of Document". In: Skare, Roswitha, Niels Windfeld Lund and Andreas Vårheim (Eds.) *A Document (Re)turn: Contributions from a Research Field in Transition*. Frankfurt: Peter Lang, 53-72.

Otlet, Paul. 1934. *Traité de Documentation*. Brussels: Editiones Mundaneum.

Otlet, Paul. 1990. *International Organization and Dissemination of Knowledge: Selected Essays*, ed. by W. Boyd Rayward. Amsterdam: Elsevier.

Pagès, Robert. 1948. "Transformations Documentaires et Milieu Culturel: Essai de Documentologie". *Review of Documentation* 15, no. 3: 53-64.

Pédauque, Roger T. 2003. *Document: Form, Sign and Medium, As Reformulated for Electronic Documents. https://archivesic.ccsd.cnrs.fr/sic_00000594. French original https://archivesic.ccsd.cnrs.fr/sic_00000511*

Pédauque, Roger T. 2006. Le Document à la Lumière du Mumérique. Caen: C&F éditions.

Pédauque, Roger T. 2007. La Redocumentarisation du Monde. Toulouse: Cépaduès.

*Proceedings from the Document Academy, 2014- http://ideaexchange.uakron.edu/docam/*Riles, Annelise (Ed.). 2006. *Documents Artifacts of Modern Knowledge*. Ann Arbor, MI: University of Michigan Press.

Schmidmaier, D. 1970. *Versuch einer Bibliographie zur bibliothekarischen Wissenschaftspädagogik im deutschsprachigen Gebiet*. 1500-1970. Freiberg Bergakademie.

Skare, Roswitha. 2009. "Complementarity: a Concept for Document Analysis?" *Journal of Documentation* 65, no. 5, 834-840.

Skare, Roswitha, Niels Windfeld Lund and Andreas Vårheim. (Eds.) 2007. *A Document (Re)turn: Contributions from a Research Field in Transition*. Frankfurt: Peter Lang.

Soto, Hernando de. 2000. *The Mystery of Capitalism: Why Capitalism Triumphs in the West and Fails Everywhere Else*. New York: Basic Books.

Stibic, Vladimír. 1982. *Tools of the Mind: Techniques and Methods of Intellectual Work*. Amsterdam: North-Holland Publishing.

Taube, Mortimer. 1952. "Special Librarianship and Documentation". *American Documentation* 3, no. 3: 166-167.

Taylor, Diana. 2003. *The Archive and the Repertoire Performing Cultural Memory in the Americas*. Durham Duke University Press.

Tricot, André, Gilles Sahut and Julie Lemarié. 2016. *Le Document Communication et Mémoire*. Louvain-la-Neuve: De Boeck - ADBS.

Wang Lin and Michael K. Buckland. 2016. "From Fief to Clan: Boisot's Information Space Model as a Documentary Theory for Cultural and Institutional Analysis". *Proceedings from the Document Academy* 3, Issue 2 http://ideaexchange.uakron.edu/docam/vol3/iss2/10

White, Howard D. 2010. Relevance in Theory. In: Marcia J. Bates and Mary Niles Maack (Eds.), *Encyclopedia of Library and Information Sciences*, 3rd ed. Boca Raton, FL CRC Press, Vol. VI, 4498-4511.

Wikipedia. 2017a. Allegorical interpretation of the Bible.
      https://en.wikipedia.org/wiki/Allegorical_interpretation_of_the_Bible Visited Feb 11,
      2017.

Wikipedia. 2017b. Douglas Engelbart. https://en.wikipedia.org/wiki/Douglas_Engelbart Visited
      Oct 6, 2017.

Wilson, Deirdre and Dan Sperber. 2012. *Meaning and Relevance*. Cambridge: Cambridge
      University Press.

Wilson, Patrick. 1983. *Second-Hand Knowledge: An Inquiry into Cognitive Authority*. Westport,
      CT: Greenwood Press.

Zerubavel, Eviatar. 1997. *Social Mindscapes: An Invitation to Cognitive Sociology*. Cambridge,
      MA: Harvard University Press.