

The ontology of documents, revisited*

Jonathan Furner

June 12, 2019

[**Slide 1**] I'd like to thank Fidélia and the organizers of DOCAM 2019 for inviting me to give this talk. It's an honor to be here in Toulon and to have this opportunity to start a conversation about what I'm calling the ontology of documents.

The title of the talk is “The ontology of documents, revisited,” which is gesturing towards the fact that I'm far from the first to give a talk on that topic. The philosopher Barry Smith has been talking about the ontology of documents since at least 2005, and part of my goal today is to shine the DOCAM lamp on his work.

[**Slide 2**] The talk is divided into 3 sections.

First of all, building from Michael Buckland's well-known paper “What is a document?” I'm going to present a brief survey of definitions of “document” from the last century or so. My conclusion from this will be that those definitions which most accurately reflect the ways in which the term “document” is used in practice are typically *compound* definitions, consisting of two or three elements. Each part refers to a different mode or function of documents: document-as-carrier or *medium*, document-as-text or *message*, and document-as-content or *meaning*. This is because documents are complex objects, not simple ones.

The second section of the talk introduces the idea of *category theory*, a branch of the philosophical subfield of ontology, whose contributors work towards the identification of the most fundamental categories of things that exist (or could possibly exist) in the world. One celebrated contributor to category theory is the philosopher E. J. Lowe, and I'm going to look at his so-called four-category ontology with a view to locating documents' place in it. As we'll see, this isn't as easy as it might initially appear to be, but my tentative conclusion is that documents are universals, not particulars. (And

*Keynote talk given at DOCAM 2019: Of Documents and Data, the 2019 annual meeting of the Document Academy (Toulon, France, June 12–14, 2019).

I just want to note from the outset that I'm going to be using that term "universal" in the narrow philosophical sense, referring to a metaphysical category, not a linguistic or cultural one. Neither is this an argument that the *concept of "document"* is a universal in the philosophical sense. That much I'm assuming to be obvious. Rather, it is an argument that *Moby-Dick* and every one of the trillions of documents that are produced daily are each universals.)

Thirdly, I'm going to switch from consideration of Lowe's work in ontology to look at that of philosopher Barry Smith, who has written specifically about the ontology of documents. At first it might seem as if Smith is working with a narrower definition of "document" than we are used to in library and information science, but I argue that ultimately we may have much to learn by taking Smith's approach, one takeaway being that what Smith calls document acts are analogous to speech acts and should be viewed as events or occurrents as opposed to objects or continuants, but another, more importantly, being that all documents, not just the ones that are involved in the kinds of acts that Smith identifies as *declarations*, are *creative* in the special sense that they are generative of quasi-abstract entities of the kind that collectively comprise social reality.

The goal of all of this though is not to draw definite conclusions, but to contribute to a solidification of connections between LIS and literatures that might not previously have loomed large on our collective radar, and to spark conversations about this material, so I'll be more than happy if I manage to do that, and I hope you'll be somewhat satisfied too.

1 Definitions of "document"

[Slide 3] Obviously the initial touchstone when it comes to any discussion of definitions of "document" is Michael Buckland's seminal and highly-cited 1997 paper. [Slide 4] This will be so familiar to everyone more than 20 years later that I hesitate to give more than the most cursory review, but I do think it's worth reminding ourselves what Buckland's goal was with this article and what he consequently chose to leave out.

Buckland explicitly poses the question "What is a document?" in the context of a *historical* discussion of the limits of *documentation*, as that activity was pursued and developed in the first half of the twentieth century. He asserts (p. 804) that such discussion is still "relevant to the clarification of the nature and scope of information systems," given the way documentation has developed in the second half of the twentieth century, but (apart from

brief consideration of “contemporary” definitions drawing from semiotics) restricts his survey of definitions of “document” to those emerging from the documentation movement prior to the mid-1960s.

[Slide 5] So Buckland begins by considering the oeuvre of Paul Otlet (1868–1944), the Belgian visionary well-known to all of us, who with Henri La Fontaine (1854–1943) founded the Institut international de bibliographie in 1895, which in 1931 became the Institut and in 1937 the Fédération internationale de documentation, and who wrote the *Traité de documentation*, published in 1934. [Slide 6] For Otlet, says Buckland, the category of “document” includes not just graphic and written records (i.e., representations of ideas or of objects) but also the objects themselves—“if you are informed by observation of them” (Buckland, p. 805). For example: “natural objects, artifacts, objects bearing traces of human activity (such as archaeological finds), explanatory models, educational games, and works of art” (Buckland, p. 805)—that is, objects “not intended as communication.” (p. 807) Comparing Otlet’s ideas to some of those promoted by modern-day cultural anthropologists and museologists, Buckland further quotes Otlet: “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, translated in Otlet 1990, cited by Buckland, p. 807)

[Slide 7] Here are some further definitions of document cited by Buckland, all except the last focusing on the supposed materiality of documents:

1935: Schürmeyer—a document is “any **material** basis for extending our knowledge which is available for study or comparison” (Schürmeyer translated by Buckland, p. 805).

1937: Institut international de coopération intellectuelle / Union française des organismes de documentation—a document is “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.” (cited by Buckland, p. 805).

1942: Donker-Duyvis—a document is “the **repository** of an expressed thought” (Donker-Duyvis 1942, translated Voorhoeve 1964, cited by Buckland, p. 806).

1951: Briet—a document is “evidence in support of a fact”; i.e., “any physical or symbolic **sign**, preserved or recorded, intended to represent, to reconstruct, or to demonstrate a physical or conceptual phenomenon” (Briet translated by Buckland, p. 806).

The last is [Slide 8] Suzanne Briet (1894–1989), of course—the French librarian [Slide 9] who co-founded the Union française des organismes de documentation in 1931 and [Slide 10] published *Qu'est-ce que la documentation?* in 1951, and who is celebrated in library schools around the world for her recognition that a photo of a star, a stone in a museum, an antelope in a zoo—all have “become physical evidence being used by those who study” them (p. 806) and therefore can be considered to be documents [Slide 11].

Especially since the publication of Ron Day’s masterful translation and analysis of *Qu'est-ce que la documentation?* in 2006, much has been written about Briet’s supposedly structuralist and proto-semiotic approach to the definition of “document.” ([Slide 12]. My own take on it is noted here.)

[Slide 13] Buckland infers from Briet four “rules for determining when an object has become a document”: 1. the object must be material (this despite the translation of Briet: “any physical *or symbolic* sign ...”; emphasis added); 2. it must have been someone’s intention that the object is to be treated as evidence; 3. the object must have been processed in some way; and 4. the object must be perceived as a document. Buckland then cites Day as specifying 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” (p. 806).

In contrast to the emphasis on the materiality of documents that pervades his prior discussion, Buckland’s conclusions include (p. 808) the identification of an “evolving” notion of “document” that has “increasingly emphasized whatever function[s] as a document rather than traditional physical forms of documents.” Buckland sees (p. 804) a move from a traditional concern with “text and text-like records (e.g., names, numbers, and alphanumeric codes)” to “any phenomena that someone may wish to observe: Events, processes, images, and objects as well as texts,” and remarks (p. 808) that “The shift to digital technology would seem to make this distinction even more important.”

[Slide 14] Buckland never intended to provide comprehensive coverage, even of the time period to which he limited himself. A few of the definitions that didn’t make it into his survey are listed here.

1907: Institut international de bibliographie (trans. Weitenkampf 1908)—“anything which represents or expresses, by the aid of any **signs** whatever (writing, image, diagram, symbols), an object, a fact or an impression.”

1943: American Library Association / Thompson—“Any written, printed, or otherwise recorded item or **physical** object that may serve as evidence of a transaction.”

1956: Ranganathan—“Record—made on more less flat surface or on

surface admitting of being spread flat when required, made of paper or other **material**, fit for easy handling, transport across space and preservation through time—of **thought** created by mind and expressed in language or symbols or in any other mode, and/or of natural or social phenomena made directly by instrument without being passed through human mind and woven into thought created and expressed by it.” [Slide 15] Buckland did of course cite the famous Indian librarian, but the work he chose was from a slightly later date.

[Slide 16] Some further definitions from American sources are listed here:

1956: Mack & Taylor—“A single piece of written or printed **matter** which furnishes evidence or information upon any subject.”

1957: Perry & Kent—“An arbitrary unit of recorded knowledge which furnishes information upon a subject. A graphic record or group of such records which are **physically** bound together or otherwise contained or attached so that it may be recognized as a single object. Examples of documents are books, reports, letters, films, photographs, and tape recordings.”

1960: Wagner—“Any recorded information **regardless of its physical form** or characteristics, and includes, but is not limited to, the following: (1) all written material, whether handwritten, printed, or typed; (2) all painted, drawn or engraved material; (3) all sound or voice recordings; (4) all printed photographs and exposed or printed film, still and motion picture; and (5) all reproductions of the foregoing, by whatever process reproduced.”

[Slide 17] Document-as-*medium*. Over the last century, various suppliers of definitions of “document” have chosen to emphasize the supposed physical, material, or concrete nature of documents. The first of Buckland’s (1997) four “rules for determining when an object has become a document,” inferred from analysis of Briet’s (1951) discussion, points to materiality as a necessary condition. (Buckland’s translation of Briet’s definition—which begins “any physical *or symbolic* sign ...” (emphasis added)—might seem to contradict this inference, but other authorities cited by Buckland, including Schürmeyer 1935 and IIC / UFOD 1937, certainly lie in the explicitly materialist camp.) We might call definitions of this kind definitions of document-as-medium, since the idea they promote is of documents as media, vehicles, or channels, for the storage and/or carrying of messages.

Document-as-*message*. At the same time, other definitions (including Briet’s) have been constructed so as to emphasize a different kind of essence—not documents’ materiality, but instead their informative, evidentiary, or signifying quality. We might call definitions of this kind definitions of document-as-message, since they represent documents as aggregations of

signs (i.e., messages or texts), for the expression and/or transmission of meanings.

Document-as-meaning. Thirdly, yet other definitions focus neither on documents' materiality nor on their signhood—i.e., not on documents as *bearers of meaning*, physical or otherwise—but on their status as *meanings in themselves*. These are definitions of document-as-meaning.

Compound definitions are those that simultaneously assign to documents two or three of the essential qualities of materiality, signhood, and meaninghood (where those latter terms stand for “being a sign” and “being meaning,” respectively). For example, Wersig & Neveling (1976) define “document” as “A unit consisting of a data medium, the data recorded on it, and the meaning assigned to the data.”

[**Slide 18**] The next few slides present, in chronological order, the highlights from a list of more than 25 definitions extracted from glossaries, dictionaries, standards, and other literature dating from 1966 through 2019.

1970: NATO Advisory Group for Aerospace Research and Development / Stolk—“a record of data, or a concept, **in any form** from which information can be derived, e.g. a page containing data, a graphic representation, a tape recording, or a book.”

1971; 1977: Harrod (3rd & 4th eds.)—“A **work** recorded in language or symbols, or by other means.”

1974: Society of American Archivists / Evans et al.—“Recorded information **regardless of medium** or characteristics.”

[**Slide 19**] 1976: BS 5408—“A **combination** of a **medium** and the information recorded on or in it, which can be used for consultation, study or evidence.”

1976: Buchanan—“Generic term for the information-bearing **media** handled by librarians – books, serials, sound recordings, films, illustrations etc.”

1976: Unesco / Wersig & Neveling—“A **unit** consisting of a data **medium**, the data recorded on it and the **meaning** assigned to the data.”

[**Slide 20**] 1983: ISO 5127-1—“recorded information which can be treated as a unit in a documentation process”

1983; 2013: American Library Association / Young (2nd ed.); Levine-Clark & Carter (“4th” ed.)—“A **physical** entity of any substance on which is recorded all or part of a **work** or multiple works. Documents include books and booklike materials, printed sheets, graphics, manuscripts, sound recordings, video recordings, motion pictures, and machine-readable data files.”

1984; 1988: International Council on Archives / Walne (1st & 2nd

eds.)—“A **combination** of a **medium** and the information recorded on or in it, which may be used as evidence or for consultation.”

[Slide 21] 1987; 1990; 1995; 2000; 2005: Harrod (6th–10th eds.)—“A record which conveys information; originally an inscribed or written record, but now considered to include **any form** of information—graphic, acoustic, alphanumeric, etc. (e.g. maps, manuscripts, tape, videotapes, computer software).”

2000: Wellisch—“A **medium** on or in which a **message** is encoded; thus, the **combination** of medium and message. The term applies not only to objects written or printed on paper or on microforms (for example, books, periodicals, maps, diagrams, tables, and illustrations) but also to non-print media (for example, artistic works, audio and video recordings, films, machine-readable records, and multimedia) and, by extension, to naturally occurring or humanly made objects intended to convey information (for example, zoo animals, plants in botanical gardens, museum collections of hand tools, etc.).”

[Slide 22] 2001: ISO 5127—“recorded information or **material** object which can be treated as a unit in a documentation process.”

2003: Feather & Sturges (2nd ed.)—“A record that contains information content. In common usage it still normally means a piece of paper with words or graphics on it. In library and information work, the term is however used to mean any information-carrying **medium**, regardless of format. Thus books, manuscripts, videotapes and computer files and databases are all regarded as documents.”

[Slide 23] 2004: Reitz—“A generic term for a **physical** entity consisting of any substance on which is recorded all or a portion of one or more **works** for the purpose of conveying or preserving knowledge. In the words of the communication theorist Marshall McLuhan, a document is the ‘**medium**’ in which a ‘**message**’ (information) is communicated. Document formats include manuscripts, print publications (books, pamphlets, periodicals, reports, maps, prints, etc.), microforms, nonprint media, electronic resources, etc. ... Also, any form printed on paper, once it has been filled in, especially one that has legal significance or is supplied by a government agency, for example, an application for copyright protection. ...”

In the paper accompanying this talk, I go ahead and list all the definitions I found, together with a coding of each of the definitions according to its membership of one or more of the three categories, document-as-medium, -as-message, and -as-meaning.

No particular *synchronic* trend is apparent in this data, but what *is* apparent is (a) the wide variety of definitions, and of combinations of category

memberships, and (b) the fact that such combinations, that is, compound definitions, are more common than single ones. I take this data as evidence for concluding that documents are (or, at least, are typically considered to be) complex objects rather than simple ones.

[Slide 24] The three-part compound definition corresponds partially to the four-entity data model that lies at the heart of IFLA's Functional Requirements for Bibliographic Records (FRBR; 1998). This model distinguishes between works (W), expressions (E) of those works, manifestations (M) of those expressions, and items (I; i.e., copies of those manifestations). Items are the only physical entities recognized in this model; each item can be viewed as the medium for (or carrier of) a given expression or aggregation of expressions, which in turn can be viewed as the message(s) representing a given work or aggregation of works. On the face of it, at least two separate conceptions of "document" can be derived from the WEMI model: one that treats documents simply as items, i.e., as physical media; and a second that conceives of documents as complex entities that exist simultaneously as material, signifying, and meaningful things, i.e., as messages (expressions) and meanings (works) as well as as media (items).

[Slide 25] As a possible way of deciding between these two alternatives, we might consider that, in the course of a discussion of the scope of FRBR (p. 8 of the final report, 1998), the word "document(s)" is used eight times as a synonym for "information resource":

"... [U]sers may make use of bibliographic records for a variety of purposes ...: to determine what information resources exist ...; to verify the existence and/or availability of a particular **document** ...; to identify a source ... from which a **document** can be obtained ...; to select a **document** or group of **documents** that will serve the information needs of the user;

"... [T]he functional requirements for bibliographic records are defined in relation to the following generic tasks that are performed by users when ... making use of ... library catalogues: using the data to find materials that correspond to the user's stated search criteria (e.g., in the context of a search for all **documents** on a given subject ...); using the data retrieved to identify an entity (e.g., to confirm that the **document** described in a record corresponds to the **document** sought by the user ...); using the data to select an entity that is appropriate to the user's needs ...; using the data in order to acquire or obtain access to the entity described (e.g., ... to access online an electronic **document** ...)."

Even though "document" is not subsequently used in the rest of the final FRBR report, the clear implication is that, in the FRBR world-view, docu-

ments are the sorts of things that are sought, found, selected, and acquired, as a result of judgments made by catalog users as to the relevance of those things, given users' needs and wants. Such judgments are made on the basis of assessments of documents' formats and contents—i.e., on the basis of evaluation of the qualities of documents as media, as messages, and as (aggregations of) meanings. On this reading, documents are not to be conceived primarily as physical items that *have* the properties of instantiating given manifestations, given expressions, and given works; rather, any given document *is*, simultaneously, a medium, a message, and a meaning.

2 Lowe's category theory

[Slide 26] So much for an empirical survey of the various kinds of definitions of “document” that have been suggested over the years. The results raise a question which demands an ontological approach: in other words, it requires some input from the philosophical subfield of ontology, the study of the nature of being.

[Slide 27] One of the tasks of ontology that has been deemed more or less important since at least the time of Aristotle (384–322 BCE) is the identification of the “highest,” “topmost,” or most general categories or kinds of things that exist in the world. Some ontologists have established systems of top-level categories that are hierarchical in structure, with one category containing all things at the very top, divided into a small number of sub-categories, each of which is subdivided into a small number of sub-sub-categories, and so on. Typically in such structures, the sub-categories at any given level are both exhaustive and exclusive, so that any individual thing is a member of one and only one sub-category at that level. The image here, for example, depicts the top-level structure proposed by the American philosopher Roderick Chisholm (1916–99) in his *A Realistic Theory of Categories* from 1996.

[Slide 28] One of the most well-known contemporary top-level ontologies is that devised by the British philosopher Jonathan Lowe (1950–2014), who wrote as E. J. Lowe and who was Professor of Philosophy for many years at Durham University in England. [Slide 29] Lowe's system, which he promoted as a means of understanding the foundations of natural science, rests on three basic binary distinctions. Lowe distinguishes between universals and particulars, between substances and properties, and between abstracta and concreta, in arriving at the structure depicted here. The diagram as presented here is lifted straight from one of Lowe's earlier publications. He

would subsequently make some changes in his use of terminology, and a few slight corrections to this diagram are needed here and there: [**Slide 30**] Properties would be merged with Relations to form a category of Attributes; the place of Relations would be taken by Kinds; and Tropes would be replaced by Modes).

Questions about whether or not these distinctions may be sustained, and if they can, how that may be done, have been among the most hotly debated in metaphysics for more than two thousand years and it's certainly not my goal to attempt to survey answers to those questions today. What I'm going to do instead is very briefly to characterize the distinctions that Lowe makes.

[**Slide 31**] Firstly, universals vs. particulars. "Even in this matter," Lowe says (2003, p. 8), "there is controversy." Lowe conceives of universals as things that are repeatable, that is, as things that may be "borne" or possessed by many different particulars, at different times and places; whereas particulars are each "wholly confined to a unique space-time location and cannot 'recur' elsewhere and elsewhen" (p. 8). In other words, universals are instantiable (by particulars), and particulars are not. Examples of universals include properties such as the property of being red, and kinds such as the kind denoted by the word "apple." Examples of particulars include the apple I ate yesterday and the redness of that apple.

[**Slide 32**] Lowe's distinction between substances and properties is among particulars. It is the distinction between objects and modes (or tropes). An object is "an entity which bears properties but which is not itself borne by anything else" (p. 8), like the apple I ate yesterday; a mode or trope is a particular that is borne as a property by no more than one object (p. 9), like the redness of that apple.

[**Slide 33**] We might say objects "instantiate" kinds, "exemplify" attributes, and are "characterized" by modes. Similarly, attributes "characterize" kinds, and are "instantiated" by modes. Another way to think of attributes is as property-kinds, in parallel with the substance-kinds that are instantiated by objects. Objects are substance-instances, characterized by modes as property-instances. It has sometimes been suggested that the so-called "four-category ontology" (Lowe, 2006) depicted in the so-called "ontological square" was first proposed by [**Slide 34**] Aristotle, and on this basis Lowe and others in his camp are known as proponents of neo-Aristotelian metaphysics. ([**Slide 35**] This version of the ontological square, by the way, was lifted from a 1997 article by Barry Smith, about whom we'll hear more in Part III.)

[**Slide 36**] The third basic distinction that Lowe draws in his top-level

hierarchy is among objects (that is, among substance-instances), and it's between abstract objects and concrete objects. Concrete objects are those that exist in space-time (that is, are "datable and locatable") or at least exist in time, whereas abstract objects are those that do not. Examples of concrete objects include individual apples; examples of abstract objects (according to Lowe) include numbers, sets, and propositions. A different criterion that may not coincide exactly in its picking out of abstract objects is the capability of an object to enter into causal relations: an abstract object is one that is incapable of such interaction.

[**Slide 37**] Controversy abounds in relation to the category of abstract object. For example, Lowe identifies propositions as a canonical sub-category of abstract object. But in what sense are propositions conceivable as particulars (that is, non-instantiable) rather than as universals instantiated by sentences (expressions of propositions) and in turn by utterances of those sentences in physical (spoken or written) form? Similarly, in what sense are works—often conceived as aggregations of propositions—comprehensible if not as universals instantiated by copies of those works in physical form?

[**Slide 38**] Another bone of possible contention has to do with the category of concrete object. Two sub-categories of concrete object highlighted by Lowe (p. 5) are masses, or material bodies, and living organisms:

"Entities belonging to these two categories have quite different existence- and identity-conditions, because a living organism, being the kind of thing that is by its very nature capable of undergoing growth and metabolic processes, can survive a change of its constituent matter in a way that a mere mass of matter cannot. A mere mass, being nothing but an aggregate of material particles, cannot survive the loss or exchange of any of those particles, any more than a set can undergo a change of its members. As a consequence, it is impossible to identify a living organism with the mass of matter which constitutes it at any given stage of its existence, for it is constituted by different masses at different stages."

Lowe does not clearly establish what other sub-categories of concrete object there are. The category of artifacts is one obvious candidate, but Lowe has little to say about artifacts in general or sub-categories of artifact more specifically, which means we're forced to speculate a little about where in his scheme certain entities might fit. In a 2014 paper entitled "How real are artefacts and artefact kinds?" Lowe distinguishes between utensils and machines as sub-categories of artifacts, arguing that machines and machine kinds, like natural kinds, are fully and mind-independently real, whereas utensils and utensil kinds—things like "tables, chairs, tents, cooking pots, knives, and hammers" (p. 24)—are not. In the conclusion to this paper,

he says “I should stress that I am not urging that machines are the *only* real artefacts. I am content to allow, for instance, that *works of art* may well qualify as real artefacts too” (p. 26, emphasis in original). At the beginning of the paper, however, he had already stipulated that he was “setting aside here putative examples of *abstract* artefacts, such as musical scores, conceived as *types* rather than tokens” (p. 18, emphasis in original) thus leaving tantalizingly open the question as to whether works of art should be counted as particulars at all.

[Slide 39] The important question for the would-be ontologist of *documents*, then, is the question of where in such a system of categories documents fit. Are documents universals or particulars? substances or properties? and so on? It’s a question that’s not as easily answered as it might at first seem. We might have an inkling of the difficulties now that we’ve conducted our survey of definitions of “document,” and especially now that we’ve distinguished between the ideas of document-as-medium, document-as-message, and document-as-meaning.

In particular, if we cleave to a *compound* definition, we might expect to have to do some extra work in situating documents among what are typically conceived as exclusive categories. And if it turns out that the concept of “document” is too “complex” for easy placement in a top-level ontology, so be it; the ontology must be revised to accommodate our concept of the thing, not vice versa.

One way to proceed is to start at the top of Lowe’s hierarchy, and attempt to justify our choice of placement of “document” on successive branches.

Beginning at Level 0, as it were, the first question is, Are documents things or not? The answer is, Yes. Since the intention is that *all* things fall in the top category of “things,” documents should be treated as things. So far so good.

At the next level down, Level 1, the question is, Are documents universals or particulars? Almost immediately, we run into a problem. On the one hand, if we consider the document-as-medium option, it seems to be fairly clear that documents are particulars (that is, they are non-instantiable). On the other hand, if we take any of the other views (simple or compound) on the nature of documents as suggested by the survey, we implicitly commit to a conception of documents as instantiable, just as (some, but not necessarily Lowe, would argue) works and propositions are.

Let’s continue, for the time being, on the assumption that the document-as-medium option is the more attractive. In that case, documents are particulars. At the next level down, then, Level 2, the question becomes, Are documents objects (substance-instances) or modes/tropes (property-instances)?

The simple answer is that, since documents are the bearers of properties, and are not borne by anything else, they are clearly objects, not modes. Next!

At the next level down, Level 3, the question is, Are documents abstract or concrete? Even for the document-as-medium conception, there is uncertainty here created by the absence of artifacts among the sub-categories of Lowe's concrete object. But this is probably a limitation of the presentation of Lowe's hierarchy and a reason to revise that presentation rather than a reason to place documents-as-media in any category other than concrete object. So far so simple.

If we make a retreat from our choice of document-as-medium, and consider how to handle documents-as-messages and documents-as-meanings, as well as documents-as-combinations, we need to go back to Level 1, and ask again, Are documents universals or particulars? In these cases, the conception is of documents as instantiable—that is, documents as texts or works that exist as types rather than as tokens. In the absence of a motivation to apply any theory that differentiates between the type-token distinction and the universal-particular distinction, our decision should be to consider documents as universals.

Moving down to Level 2, the question becomes, Are documents substance-kinds or property-kinds (attributes)? Since documents are characterizable, the simple answer is substance-kinds.

[Slide 40] So we have a situation where on the one hand, documents “live” in a new sub-category of concrete objects, possibly called artifacts, where the challenge will be to identify the qualities that distinguish documentary artifacts from non-documentary artifacts (if there are such things as the latter); and on the other hand, “document” is placed in the category of “Kinds.” There is actually a third possibility not suggested by our navigation through Lowe's hierarchy, but which is nevertheless suggested by Lowe's characterization of the category of abstract objects as including not just the sub-categories of numbers and sets, but also propositions. If works are considered to be aggregations of propositions, then it might seem that works should also be placed here; from work, it's a short step to document-as-meaning, and thus to thinking of documents as abstract objects.

Ideally we would use different words to refer to these different concepts. But we don't. We use a single word, “document,” interchangeably in different contexts to mean different things. Perhaps this is no bad thing. It keeps us on our toes. But I want to suggest that much of the time, many people use the word “document” to mean something that's a universal not a particular—something that's not *necessarily* material. And I think that's

important for our understanding of documents, and for our ideas as to where document theory could or even should be going.

In essence, my argument is an empirical one about the use of language. What do we talk about when we talk about documents? Much of the time, the properties of documents that we're most interested in are properties of documents-as-meaning, or if you like, documents-as-works, documents-as-universals, documents-as-types that are multiply instantiated by physical tokens. Of course it's important to recognize the existence of and to understand the nature of the type/token relationship, but that doesn't necessarily mean that the sole or even the primary focus of document theory should be on the materiality of document tokens. Instead, or at least in addition, understanding the universality of document types should be high on our agenda.

We've reached the end of Part II. In the third and final part of the talk, I'd like briefly to discuss what I take to be the most substantive contribution to the ontology of documents of the last few years.

3 Smith's ontology of documents

[Slide 41] The philosopher Barry Smith published a short conference paper called "The ontology of documents" in 2011, following up on a presentation on a similar topic from 2005. [Slide 42] In these and several related papers on his so-called "theory of document acts," Smith has developed an account of the status of documents in the context of the picture of social reality painted by [Slide 43] fellow philosopher John Searle over a period of several decades.

[Slide 44] This work is not really about categories per se, but is ontological in the sense that it explains how documents—or more precisely *document acts* like signing a document, filling it in, delivering it, or archiving it—have the effect of bringing new entities into existence, that is, how document acts have ontological consequences.

[Slide 45] Smith builds on John Searle's ideas about speech acts—the things we do with words. Searle explains how certain kinds of speech acts—the ones he calls "declarations"—can bring about changes in the ontology of social reality, and Smith similarly describes how certain kinds of document acts—generally speaking, the things we do *with* rather than *to* documents—bring into existence not just *physical* entities like document tokens, and document-related artifacts like filing systems, but also document-related social practices and *quasi-abstract* entities ("at one and the same time *subject*

to historical changes yet not made of physical parts,” Smith 2011, p. 2, emphasis in original), especially in the realms of commerce, law, and government. [Slide 46] These quasi-abstract entities are things like organizations, contracts, laws, money, rights, obligations, identities, claims, privileges, corporations, capital, permissions, debts, trusts ... that is, they are entities that form vitally important parts of social reality.

Smith argues that we need to pay more attention to document acts than Searle does. His primary reason for doing so is to address issues relating to the “anchorage” of digital documents to the people who created them, by implementing systems for the certification of authenticity that are on a par with signatures and fingerprinting for physical documents.

But in the course of setting this up, Smith also makes important contributions to our understanding of the basic ontological categories to which documents belong.

[Slide 47] Firstly, Smith points out (as others have done before and since) how Searle’s categorization of speech acts may be applied to documents. Thus we may distinguish among documents that are representative, directive, commissive, expressive, and declarative. Representatives “commit the [writer] ... to the truth of ... expressed proposition[s]”; directives “attempt ... to get the [reader] to do something”; commissives “commit the [writer] ... to some future course of action”; expressives “express the [writer’s] psychological state ... about a state of affairs”; and declarations “bring about correspondence between ... propositional content and reality” (Searle 1975, pp. 354-361).

[Slide 48] Smith is most concerned with declarations, those which are most clearly generative of new entities. Smith uses the term “creative” rather than generative, with the somewhat counter-intuitive result that he classes books (fiction and non-fiction), journal articles, maps, artworks as “non-creative”; on the other hand, certificates, contracts, receipts, banknotes, licenses, agreements, filled-in forms, passports, diplomas, medical records, meeting minutes, etc., all have “creative power” in social reality.

[Slide 49] Secondly, Smith follows American philosopher Nelson Goodman (1906–98) in distinguishing those documents that are autographic, and those that are allographic. [Slide 50] For Goodman (*Languages of Art*), a work of art is autographic “if and only if the distinction between original and forgery of it is significant; or better, if and only if even the most exact duplication of it does not thereby count as genuine.” (Goodman 1976, p. 113). [Slide 51] Painting, sculpture, and architecture are autographic; music, photography, and literature are allographic.

The autographic/allographic distinction seems to correspond at least

roughly to the distinction between documents that exist as both types and tokens, and those that do not tokenize a type. Briet’s photograph of a star? Allographic. The stone in a museum, and the antelope in a zoo? Autographic. Many of the “creative” document-types listed above? Autographic. Whether the type/token distinction itself corresponds to the universal/particular distinction is a complex matter that might be better left for another day, although clearly any conclusions will be significant for our decision-making when it comes to situating documents in a top-level hierarchy of categories like Lowe’s.

[Slide 52] Thirdly, and perhaps also significantly for that decision-making, Smith seems to conceive of document acts, including document production acts, like Searle’s speech acts, as events or occurrents, notwithstanding that the physical records that are among the products of such acts are definitively classed by Smith as continuants (eliding the type/token distinction). Scope for further sub-categorization of concrete objects?

[Slide 53] A more general question raised by Smith’s analysis is, What might be the more productive route for Document Academy-style document theory to follow? On the one hand, we might imagine a future document theory that commits wholeheartedly to the distinction drawn by Smith between creative and non-creative documents, carving out a subfield that focuses on the former and on kinds of issues identified by Smith as critical for a digital social reality whose effective and efficient organization depends so much on reliable authentication of autographic creative documents. In this way we may contribute to the kind of “scientific understanding” that Smith says is necessary for arriving at an “intelligent appreciation of the changes in social reality that are being effected through the trillions of documents being created daily in the digital realm” (Smith 2014).

On the other hand, might it be productive to extend the notion of declarative documents’ creative or generative power so that all documents, including those that Smith identifies as non-creative but that are traditionally the main concern of document theory, are considered to be creative in some respect and/or to some degree. Smith enumerates some of the kinds of things one can do *to* a document, such as sign it, fill it in, register it, and archive it. It is surely a short step to take to consider that some other such document acts include finding it, identifying it, selecting it, and obtaining access to it (just to choose those suggested by the FRBR final report), as well as organizing it, classifying it, and indexing it, and reading it, interpreting it, citing it, and using it, in many and various ways. [Slide 54] Similarly, the products of such acts include quasi-abstract entities of many and varied kinds, including metadata, bibliographies, catalogs, result-sets, recommendations,

rankings, metrics, and networks, to name just a few.

[**Slide 55**] And there I'll stop, with a reminder of my three conclusions, [**Slide 56**] and a list [**Slide 57**] of references. [**Slide 58**] I'd like to thank my colleagues Greg Leazer and Julie Park for their helpful comments and thank you very much for listening. [**Slide 59**]