

Documents as Timed Abstract Objects

Claus Huitfeldt

Associate Professor

Department of Philosophy, University of Bergen, Norway

`<claus.huitfeldt@uib.no>`

Fabio Vitali

Department of Computer Science, University of Bologna, Bologna, Italy

`<fabio@cs.unibo.it >`

Silvio Peroni

Department of Computer Science, University of Bologna, Bologna, Italy

`<speroni@cs.unibo.it >`

Balisage: The Markup Conference 2012

August 7 - 10, 2012

Copyright © 2012 by the authors

How to cite this paper

Huitfeldt, Claus, Fabio Vitali and Silvio Peroni. "Documents as Timed Abstract Objects." Presented at Balisage: The Markup Conference 2012, Montréal, Canada, August 7 - 10, 2012. In *Proceedings of Balisage: The Markup Conference 2012*. Balisage Series on Markup Technologies, vol. 8 (2012). doi:10.4242/BalisageVol8.Huitfeldt01.

Abstract

At Balisage 2009 and 2010 Renear and Wickett discussed problems in reconciling the view that documents are abstract objects with the fact that they can undergo change. In this paper we present an account of documents which we believe is quite common, but which was not discussed by Renear and Wickett.

According to this account documents are indeed abstract objects, but this is easily reconciled with the fact that they are created and can undergo change. We then point to a similarity between this account and the notion of so-called space-time slices. We argue that the proposed account of documents as timed abstract objects may be subject to the same kind of criticism that has been raised against the notion of space-time slices.

We believe that our account fares no worse than the other accounts given of documents as abstract objects. But it still fails, and we remain agnostic about the ontological status of documents and their relation to abstract objects, as well as about the nature of abstract objects. We conclude that either documents are not (or not related to) abstract objects, or they are (or are related to) abstract objects of a kind which does not correspond to the standard definition of what an abstract object is.

Table of Contents

Background

The Inconsistent Triad

Responses to the Inconsistency

Background

Cultural objects like literary or musical works and other repeatable forms of expression occupy an ambiguous and philosophically interesting position between the concrete and the abstract. On the one hand, they become known to us only in the form of concrete physical objects like books, scores or performances. On the other hand, it seems unnatural to identify the works with their manifestations: A literary work is not identical with any sum of physical manifestations, and a musical work is identical neither with its score nor with its performances.

Questions concerning the ontological status of literary or musical works fall within the wider domain of questions concerning the ontological status of other cultural or artistic expressions such as for example sculpture or painting, and also within the even wider domain of questions concerning the status of cultural artifacts in general. Our discussion in this paper is limited to the realm of writing.

Following Goodman [Goodman 1976], we may distinguish what may be called "unique" from "multiple" art forms by appeal to his definitions of autographic and allographic works. . Roughly, and limiting ourselves to modern Western culture, literary and musical works are based on notations and are therefore allographic, whereas sculptures and paintings are non-notational, and therefore autographic. No copy of Mona Lisa or replica of David is an authentic manifestation of Mona Lisa or David, whereas any edition of Hamlet or performance of Bolero is a manifestation of Hamlet and Bolero, respectively. In the former case, the authenticity of a manifestation is a question of physical identity. In the latter case, no question of authenticity in terms of physical identity applies. Whether a copy is a manifestation of a literary or musical work is only a question of correctness of reproduction within certain constraints.

Thus, the identity criteria for autographic works are similar to the identity criteria for physical objects. The identity criteria for allographic works are slightly more difficult to specify. What is it that makes two physical objects manifestations of the same document? ^[1] And, considering the fact that a physical manifestation of a document may undergo change: How do we decide whether or not it is still a manifestation of the same document?

One way of accounting for the identity of documents through multiplicity and change among physical manifestations is to say that the document is an abstract object, and that its manifestations are instantiations of that abstract object. So what two physical objects must have in common in order to be manifestations of the same document is that they must both instantiate the same abstract object. ^[2]

In philosophy this kind of explanation by reference to abstract objects is perhaps contentious, but not uncommon. We observe that it is perhaps more common, and certainly regarded as less contentious, in domains outside philosophy. In the library professions and library and information science, for example, one of the currently important standards is Functional Requirements for Bibliographic Records [FRBR 98], distinguishes between bibliographic objects in terms of different abstraction levels such as item, manifestation, expression and work.

The Inconsistent Triad

Renear and Wickett point out that the view that documents are abstract objects leads to the following inconsistent triad:

- Documents are strings.
- Strings cannot be modified.
- Documents can be modified.

The authors argue that all the three statements seem to have some initial plausibility, yet any two of them implies the negation of the remaining statement.

In the example as quoted, the explicit reference is to "strings", not to abstract objects. The authors make it very clear, however, not only that they regard strings as abstract objects, but also that they believe their argument to be valid irrespective of which particular kind of abstract object documents are thought to constitute. They take care to make explicit that the inconsistency cannot be seen to trade on any ambiguity in the way that the three statements are phrased. ^[3] In [Renear & Wickett 2010] they observe that the inconsistency does not arise if the last of the three statements is read not as an existence claim, but only as a universal generalisation.

For purposes of terminology and generality, we may therefore reformulate the inconsistent triad as follows:

- 1. All documents are abstract objects
- 2. No abstract object can change.
- 3. Documents exist, and they can change ^[4] .

Responses to the Inconsistency

It is not hard to agree with Renear and Wickett that the third of the three statements, that documents exist and that they can change, has "initial plausibility", as it seems to be implied in all our talk and all our dealings with documents as objects of our everyday world. The first statement may seem less obvious, but the authors argue forcefully, and in our eyes they are right, that it is an underlying assumption in much or most of the theoretical disciplines and technologies involved in document processing and management that documents are abstract objects. The second, i.e. the claim that abstract objects cannot change, is in accordance with the standard modern definition of abstract objects as non-spatial, atemporal and causally inefficient. ^[5] We have nothing to say against the plausibility of the three statements at this stage, though we will come back to the point about abstract objects towards the end.

When the three statements form an inconsistent triad, and the inconsistency does not rest on some ambiguity, the natural question to ask is which of the statements to reject or modify. Renear and Wickett go through and argue against a considerable number of possible responses. Rejection of the statement that documents are abstract objects (in the form of "materialist" or "social object" strategies) as well as the rejection of the statement that documents can change (in the form of "new document" or "selection" strategies) are considered in Renear & Wickett 2009. Rejection of the statement that documents exist is considered in Renear & Wickett 2010. As a more acceptable alternative, they propose a "string-in-a-role" strategy, which partly denies the statement that documents can be modified and partly denies the statement that documents are abstract objects, by redefining the notion of a document. Although we do not think that all their arguments in favour of their own and against

the other strategies are equally convincing ^[6], we will not pursue these matters here, but rather point to yet another alternative.

Timed Abstract Objects

Science is not unfamiliar with the problems involved in describing time-related events and processes in physical reality in terms of time-independent mathematical concepts. A whole subfield of physics, dynamical systems, is devoted to the formal study of the mathematics of things that change. Poincaré (1854-1912) and Gibbs (1839-1903) provided the basic tools used today to study anything from the evolution of chemical reactions to the trajectories of celestial bodies. The basic approach of dynamical systems is to use functions of time to model the evolution of a point in a manifold. The set of points of the manifold represent all the potential states, the phase space, and each function draws a plot or a trajectory within the phase space. The evolution function is often defined as an iterative function that determines the state of the system in the future, given its conditions at the present. The study of the dynamical system is meant to determine the state of the system at any point in the future by solving the equations of the evolution function.

In computer science, the theory of automata studies the behavior of abstract machines that change their internal state upon receiving some input. More specifically, an automaton is an abstract machine that passes from a single initial state through several intermediate states to one of many final states (e.g., Success) depending on the input received. In this model, the automaton is a quintuple $(Q, \Sigma, \partial, q_0, F)$, where Q is the set of states (the state space) of the machine, Σ is the alphabet of the potential inputs, ∂ is a transition function from one state to another depending on elements of the alphabet Σ , q_0 is the initial state, and F is the subset of Q that constitute the acceptable final states. An automaton succeeds on input X (a string of items of the alphabet Σ) if the machine transits from the initial state through intermediate states up to a final state while receiving input X , and fails if it remains in an intermediate state.

Automata are more useful than dynamic systems in our discussion: Dynamic systems analysis is well suited for modelling change in manifolds, i.e. continuous state spaces. Automata theory, however, is better suited to deal with discrete structures, such as strings. We may regard strings as the state space of the content of a document.

On this conception, strings are still immutable concepts that do not change over time. The number 2 is an immutable object. The string “To be or not to be” is also an immutable object. The set of all strings, past, future or potential, each one of them an immutable object, is also an immutable object. We could now proceed in two different, but ultimately equivalent, directions: On the one hand we could go the dynamical system way, introducing a change function for each document. This function is associated with the ordered independent variable similar to “time”, which we will call **time*. ^[7] On the other hand, we could introduce an edit function, associated with an independent variable called “edit session”. In both cases, the function selects a specific string for each value of time t or edit session s . For instance, it will select string s_1 for time t_1 , and string s_2 for time t_2 , etc. Given a change function c or an edit function e , we can now define a document as the sequence of strings s_c that are selected for each value in time or edit session.

Still, documents, like strings, are immutable concepts that do not change over time. But documents are not strings. While not changing in time, documents change in **time*, because the function change

associates different strings with different *time points. This allows us to provide a time-dependent view of documents that is not affected by the immutableness of the basic mathematical concepts under it: we just select different strings at different *times and regard them as the state of the document at that *time.

Timed Abstract Objects and FRBR

When we refer to a document, we can mean to refer to several different objects depending on our backgrounds and social roles (e.g., reader, archivist, author, publisher, etc.). In fact, as [Tillett 2004] puts it, when we speak about a document, we may refer to a physical object in our bookcase, to a particular publication (i.e., an edition) identified through an ISBN, to its content (or part of it) identified through a DOI and, finally, to the abstract conceptualisation, i.e., «the conceptual content that underlies all of the linguistic versions, the story being told in the book, the ideas in a person's head for the book» [Tillett 2004].

The Functional Requirements for Bibliographic Records (FRBR) [IFLA 2009] is a general model, proposed by the International Federation of Library Associations and Institutions (IFLA),^[8] for the description of documents and their evolution according to the aforementioned perspectives. To describe FRBR,^[9] we start from physical objects and work upwards toward more abstract concepts.

Let us consider a physical book. We can identify some of its properties: weight, size, number of pages, color of the cover, quantity and shape of the ink on the 15th page, and so on. If we partake to a culture that recognizes the book as a cultural object, a *book, may be able to identify more properties: a title, an author, a publication date, a content, and so on. Now let us consider another book. We notice that it has the same physical properties as the first one (same weight, size, color, disposition of the ink in each page etc.) and, by interpreting the underlying *book, we notice it has the same title, author, content, etc. We need a concept to collect and describe these similarities against the fundamental difference that they are, in fact, numerically different objects. Thus we define each physical object as an Item of the same *book, and allow this *book to be perceivable as one through many different books, each having the same values for most cultural and physical properties, except for physical sameness.

Now let us consider a PDF file, which we are told is the source of the physical books seen before: it is physically completely different, being a computer structure rather than a physical object. Yet, somehow the culturally relevant concept of “content” is the same, and a number of culturally relevant related concepts (e.g., the author, the title, and so on) are identical. We have two different forms of book, but we interpret both forms as representing the same *book. We call them individual Manifestations of the same *book, and allow the *book to be perceivable as one through many different forms, each having the same values for a number of cultural properties, but many different physical properties.

Let us next consider another book, (either electronic or physical, it does not matter), which we are told is a different version of the same *book. This book is different from all the others, even in content: it may be a slight difference, a few words here and there, or a complete difference, as in having close to no similar content.^[10] Possibly, a number of cultural properties are the same: same author, same title (which does not mean the same string for the title property, but same conceptual value for the title concept), etc, and somehow we recognize that these two books are related, that they

are one, in some more abstract way than by sharing the same content. We may have some properties that have the same values, but we may also have no sameness whatsoever, except only for some culturally shared and accepted, yet vague and approximate, sameness. We call the content of the two books, and the properties that are specific of each content, two Expressions of the same *book, and allow the *book to be perceivable as one through many different contents, each different in most or all properties except for some vague but culturally accepted sameness.

We call Work whatever is left, composed of any property we agree is still associated with the *book regardless of the Item, Manifestation and Expression it is perceivable through, or nothing but the above-mentioned indistinct sameness, if no property is left. The Work is characterized by those properties that are shared by all Expressions, and in turn all Manifestations and all Items of a *book. On the other hand, each of its Expressions and Manifestations exist in a multiplicity of instances. At different points of *time we may access the same *book and find different contents, different forms, even different physical properties (a corner is torn, a stain appears on the cover, the book is physically replaced by the bookstore due to a misprint, etc.). Thus Expressions, Manifestations and Items change in *time and are, by definition, changeable and modifiable.

The FRBR model has been used in different contexts: for the publishing domain [Peroni & Shotton 2012], in open-government contexts [Cifuentes Silva et al. 2011], for the description of the musical publishing process [Raimond et al. 2007] and for the legal domain [Barabucci et al. 2009]. For instance, Akoma Ntoso^[11] – a set of simple technology-neutral XML formats for parliamentary, legislative and judiciary documents – makes wide use of the FRBR hierarchy. It re-interprets parts of the FRBR endeavours so as to clearly and unambiguously identify the core attributes that characterise legislative documents. The entity Work is associated with the concept of identity, the Expression and Manifestation layers relate respectively to the content and the format of a such a document, and Item is used to indicate the particular location in which a concrete document is available.

In [Renear & Wickett 2010] the authors explain that although «[t]he word "document" has ... many senses», their own «informal characterization of the sense ...[is]... more or less the same sense implied in the XML specification, which defines "XML Document" as a string that matches certain formal constraints [Bray et al. 2008]», and also that it «corresponds closely to the concept of an "expression" in the "Functional Requirements for Bibliographic Records" (FRBR) [FRBR 98]».

Yet on the one hand, the definition of XML documents seems more related to the Manifestation layer than to the Expression layer of FRBR: we read in Bray et al. 2008 that «an XML document may consist of one or many storage units. These are called entities; they all have content and are all (except for the document entity and the external DTD subset) identified by entity name». Thus, XML seems to be a vehicle to store document contents in a particular format by means of entities, which seems closer to the FRBR definition of Manifestation than that of Expression.

On the other hand, however, the association of XML documents with only one of the various abstraction layers of FRBR is in our eyes unfortunate, be it the Expression, the Manifestation or the Item layer. We believe that FRBR defines a specific endeavour to model the time-dependency of documents, i.e., the fact that the Work is associated in time with a number of different Expressions. It is only by allowing XML documents to be considered, depending on context, either as Items, as Manifestations, as Expressions or even as Works in their own right, that the full potential of applying FRBR also to electronic documents can be exploited. ^[12]

The FRBR-aligned Bibliographic Ontology (FaBiO)^[13] provides an example of how changes to a document may be tracked. For instance, we may summarise the steps of the creation of a particular research paper (class `fabio:ResearchPaper`, a subclass of `frbr:Work`). Let us assume that this particular Work has three authors, as shown in the following excerpt (in Turtle syntax [Prud'hommeaux & Carothers 2011]):

```
:paper a fabio:ResearchPaper
; frbr:creator :author1 , :author2, :author3 .
```

The very first (incomplete) draft content of this Work (i.e., an Expression) is written by one author only (let us suppose `:author1`). In FaBiO, this particular scenario can be described by referring to a particular realisation of the Work, an individual of `fabio:ConferencePaper` (i.e., a subclass of `frbr:Expression`), as follows:

```
:paper frbr:realization :first-content-draft .
      :first-content-draft a fabio:ConferencePaper
; frbr:realizer :author1 .
```

In this way, we can record both that there are three authors of the paper and that `:author1` was responsible for writing the first draft of that paper. Similarly, the writing process can proceed with `:author2`, who wrote, for instance, a revised version of the content written previously by `:author1`:

```
:paper frbr:realization :revised-content-draft .
      :first-content-draft frbr:revision :revised-content-draft .
      :revised-content-draft a fabio:ConferencePaper
; frbr:realizer :author2 .
```

This process can be repeated until the authors produce the final version of the paper to be submitted to a venue. Has the document changed in time? If one considers each expression as a separate document (as suggested by Renear & Wickett [Renear & Wickett 2010]), then the answer is surely «no, none of the documents have changed». However, none of the previous documents record both `author1`, `author2` and `author3` as its authors. Thus, none of those documents fully represents the paper we are talking about, since it should record three people as its authors.

Everything changes if we take the Work `:paper`, rather than its various Expressions, into account. In this case, the author is clearly specified – i.e., all the people involved in the realisation of each Expression are authors of `:paper`. The Work in consideration has changed during **time*, since it has been realised in several revised expressions that differ in content and in **time* of creation.

Problems with Timed Abstract Objects

We believe that this account of documents as Timed Abstract Objects has a number of strengths, and that it gives a coherent explanation of document change as well as our practices of identifying documents at different levels of abstraction. Notwithstanding its attractiveness, however, it also has some weak points. One of its consequences is that two documents which would perhaps on other accounts be regarded as identical would have to be regarded as different if they have different histories.^[14] In other words, two series of identical changes to "the same" document at slightly different times will constitute two different documents. But perhaps this is just a minor quirk, perhaps

we simply don't have clear intuitions in such cases.

However, the fact that its change history becomes so to speak a constitutive part of the document itself also implies that documents are not really accounted for as changeable objects, but as "events". Events, though they do take time, don't change. This attempt to account for document change has the seemingly paradoxical consequence that documents do not really change at all. (A poker may exemplify change by being first cold, and then warm. But a poker which is cold at the one end and warm at the other does not exemplify change.)^[15]

It is interesting, and may be instructive, to notice the parallel to the notion of space-time slices proposed by so-called "four-dimensionalism" in ontology, and some of the criticisms of that notion. The notion of space-time slices was introduced by philosophers like Quine and Goodman in an attempt to salvage the claim that concrete physical objects (or more generally, individuals) are identical with the sum of their parts (the claim that "the whole is nothing over and above its parts"), yet can undergo change. If an object is identical to the sum of its parts, then how can it lose a part, and yet remain the same object? One favoured example in the discussion is a cat losing its tail. After the loss of its tail, the cat is still the same cat, but clearly not the same sum of parts.

The answer proposed by four-dimensionalism is that the cat, correctly considered, is not the sum of its parts at any single moment, but the sum of all its temporal as well as its spatial parts. And when the cat is considered as the sum of all its spatio-temporal parts, there is no difficulty explaining how references to the cat before and after the loss of its tail can be references to the same cat: they refer to two spatio-temporal parts of the very same cat. What we have before us at any particular moment when we consider the cat is not the whole cat, but a temporal part, a time-slice, of the cat.

One of the critics of this conception, Peter Simons, points out that on the four-dimensional account, the cat losing its tail does not really *change*: The loss of its tail is so to speak baked into the cat as such. Simons' complaint "is not that the four-dimensional ontology is inconsistent (if it is inconsistent, this is far from obvious), but that no one has begun to do the work required to make it understandable to we users of a continuant/occurrent ontology".^[16]

It may be conceded that a proposal to consider documents as Timed Abstract Objects is not quite as drastic as the analogous attempt to account for continuants in general as four-dimensional objects. But, as suggested above, we do think that the proposal has similar, counter-intuitive implications. Finally, the account may have counter-intuitive consequences when applied to the actual handling of documents in every-day contexts, especially involving digital documents. When deciding whether two document files instantiate the same document we do not normally, and would normally not be in the position to, decide whether or not they have been brought about by the same string-modifying operations at the same points of time, nor is it easy to think of examples where such information would be considered relevant.^[17]

More importantly, however, we believe that the difficulties with this and the other proposals to deal with the inconsistent triad may be rooted in the concept of abstract object presupposed by all of them. Renear and Wickett explicitly exclude from consideration "responses that deny the second assertion", i.e. (in our rephrasing) that abstract objects cannot change.

We have come to believe that as long as abstract objects are defined as atemporal, non-spatial and causally ineffective, none of the above accounts can explain the perhaps most important distinctive

feature of documents: That they can carry information, be written and read by human beings, and thus serve the transmission of knowledge, culture and value.

Admittedly, Renear and Wickett's proposal is not that documents themselves are abstract objects, but that they are abstract objects in a particular role. Similarly, on the Timed Abstract Object account the claim is not that documents themselves are abstract objects, but that they are functions from time sequences to abstract objects. On both accounts, however, documents are defined in relation to abstract objects.^[18]

The problem of explaining how we can have epistemic access to abstract objects is well known. For example, Swoyer, a realist about abstract objects, says:^[19]

Epistemology is the Achilles' heel of realism about abstracta. We are biological organisms thoroughly ensconced in the natural, spatio-temporal causal order. Abstract entities, by contrast are atemporal, non-spatial, and causally inert, so they cannot affect our senses, our brains, or our instruments for measuring and detecting.

It may very well be that atemporal, non-spatial and causally ineffective abstract objects exist. Perhaps mathematical objects like numbers or sets are of this kind (even though similar problems concerning epistemic access apply here). But we seem forced to conclude not only that documents are not themselves abstract objects, but also that any account of documents that relies on our epistemic access to abstract objects remains unsatisfactory unless the definition of abstract objects on which it relies differs from the standard one.

Conclusion

Among the considered accounts of documents in terms of abstract objects, we believe that the Timed Abstract Object account is the best. However, the failure of all accounts of documents in terms of abstract objects to explain how we can have epistemic access to them is a major drawback. The fact remains that documents exist, that they are created and can be changed by human beings, and that they can go out of existence. We remain agnostic about their ontological status, but we conclude that either documents are not abstract objects after all, or they are abstract objects of a kind which does not correspond to the standard definition of what an abstract object is.

References

- [Barabucci et al. 2009] Barabucci, G., Cervone, L., Palmirani, M., Peroni, S., Vitali, F. (2009). "Multi-layer markup and ontological structures in Akoma Ntoso". In *Proceeding of the International Workshop on AI approaches to the complexity of legal systems II* (AICOL-II). Rotterdam, The Netherlands.
- [Barabucci et al. 2010] Barabucci, G., Cervone, L., Di Iorio, A., Palmirani, M., Peroni, S., Vitali, F. (2010). "Managing semantics in XML vocabularies: an experience in the legal and legislative domain". In *Proceedings of Balisage: The Markup Conference 2010*. Montreal, Canada. doi:10.4242/BalisageVol15.Barabucci01.
- [Bray et al. 2008] Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., Yergeau, F. (2008). *Extensible Markup Language (XML) 1.0* (Fifth Edition). W3C Recommendation 26 November 2008. World Wide Web Consortium. <http://www.w3.org/TR/REC-xml>
- [Cifuentes Silva et al. 2011] Cifuentes Silva, F. A., Sifaqui, C., Labra Gayo, J. E. (2011). "Towards an architecture and adoption process for linked data technologies in open government contexts: a case study for the Library of Congress of Chile". In *Proceedings of the 7th International Conference on Semantic Systems (I-SEMANTICS 2011)*. Graz, Austria, September 7-9.

- [Dorr 2008] Dorr, Cian. "There Are No Abstract Objects" in Sider et. al 2008, pp. 32-63.
- [FRBR 98] *Functional requirements for bibliographic records : final report / IFLA Study Group on the Functional Requirements for Bibliographic Records*. — München : K.G. Saur, 1998. — viii, 136 p. — (UBCIM publications ; new series, vol. 19). — ISBN 978-3-598-11382-6.
<http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records>
- [IFLA 2009] International Federation of Library Associations and Institutions Study Group on the Functional Requirements for Bibliographic Records (2009). Functional Requirements for Bibliographic Records Final Report. International Federation of Library Associations and Institutions. http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf
- [Goodman 1976] Goodman, Nelson. *Languages of art: An approach to the theory of symbols*. Indianapolis, Cambridge: Hackett, 1976.
- [Peroni & Shotton 2012] Peroni, S., Shotton, D. (2012). *FaBiO and CiTO: ontologies for describing bibliographic resources and citations*. Submitted for publication in the Journal of Web Semantics.
- [Prud'hommeaux & Carothers 2011] Prud'hommeaux, E., Carothers G. (2011). *Turtle, Terse RDF Triple Language*. W3C Working Draft 09 August 2011, World Wide Web Consortium.
<http://www.w3.org/TR/turtle>
- [Raimond et al. 2007] Raimond, Y., Abdallah, S., Sandler, M., Giasson, F. (2007). "The Music Ontology". In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*. Vienna, Austria, September 23-27.
- [Renear & Wickett 2009] Renear, Allen H. and Karen M. Wickett *Documents Cannot Be Edited*. Presented at Balisage: The Markup Conference 2009, Montréal, Canada, August 11 - 14, 2009. In *Proceedings of Balisage: The Markup Conference 2009*. Balisage Series on Markup Technologies, vol. 3 (2009). doi:10.4242/BalisageVol3.Renear01. Downloadable from
<http://balisage.net/Proceedings/vol3/html/Renear01/BalisageVol3-Renear01.html>
- [Renear & Wickett 2010] Renear, Allen H. and Karen M. Wickett *There are No Documents*. Presented at Balisage: The Markup Conference 2010, Montréal, Canada, August 3 - 6, 2010. In *Proceedings of Balisage: The Markup Conference 2010*. Balisage Series on Markup Technologies, vol. 5 (2010). doi:10.4242/BalisageVol5.Renear01. Downloadable from
<http://balisage.net/Proceedings/vol5/html/Renear01/BalisageVol5-Renear01.html>
- [Rosen 2012] Rosen, Gideon. "Abstract Objects" *The Stanford Encyclopedia of Philosophy* (Spring 2012 Edition), Edward N. Zalta (ed.). <http://plato.stanford.edu/archives/spr2012/entries/abstract-objects>.
- [Simons 1987] Peter Simons. *Parts. A Study in Ontology*. Oxford University Press 1987.
- [Sider et. al 2008] Theodore Sider, John Hawthorne, and Dean W. Zimmerman (eds.): *Contemporary Debates in Metaphysics*. Blackwell 2008.
- [Swoyer 2008] Swoyer, Chris. "Abstract Entities" in Sider et. al 2008, pp. 11-31.
- [Tillett 2004] Tillett, B. (2004). "What is FRBR? A conceptual model for the bibliographic universe". *Library of Congress Cataloguing and Distribution service*.
<http://www.loc.gov/cds/downloads/FRBR.PDF>

^[1] So far, we have talked about musical and literary "works". From this point on, we will talk mainly about "documents": This may be slightly confusing, as many writers on this subject use the term "document" to refer to the physical aspect of the matter, while "text" or "work" refer precisely to the abstract aspect. For the sake of this discussion, however, we stick to Renear and Wickett's terminology.

^[2] At this point we have already departed completely from Goodman, who had little patience with any talk of abstract objects.

[3] In [Renear & Wickett 2009], they offer the following formalisation:

- $(x)[(isaDocument(x) \rightarrow isaString(x))]$
- $(x)[(isaString(x) \rightarrow \sim isModifiable(x))]$
- $(Ex)[(isaDocument(x) \& isModifiable(x))]$

[4] Since this is a conjunction of two statements, the inconsistent "triad" might actually be said to consist of four statements, rather than three.

[5] See for example [Rosen 2012] and [Swoyer 2008]. If everybody agrees that abstract objects cannot change and that documents can change, it may seem futile to discuss whether documents are abstract objects. However, it does not follow from these two claims alone that documents, or our conceptions of them, do not presuppose or are not in some other sense basically related to abstract objects. Moreover, the notions that abstract objects are atemporal, non-spatial, and causally inefficient, and that documents are in some sense abstract, are so deep-rooted in modern science and culture that it seems reasonable to take the discussion seriously, futile or not.

Even so, to the authors of this paper (and probably to many of its readers) it often seems awkward to use the term "object" (including "abstract object") about abstract topics of scientific discourse, where we would often be inclined to talk about concepts rather than objects. If we agree that a chair is an object and freedom is a concept, for example, it seems more plausible to liken documents, strings, and mathematical structures (say π) to freedom than to the chair. The relation between concepts and objects as technical terms in this context is a complex one, however, and far too far-reaching for this paper. For purposes of the present discussion, therefore, we try to keep with the "object" terminology, but the reader may notice that we have sometimes, when this has seemed too strained, allowed ourselves to use the term "concept".

[6] For example, some views are simply dismissed as "difficult to reconcile with a naturalistic view of the world", without further argument.

[7] One may need to make further restrictive hypothesis about the *time points, if he/she were inclined to over-analyze this theory, but for the moment we let it stand that *time may not be time but behaves like time.

[8] <http://www.ifla.org/>

[9] ... or, rather, our own understanding of the FRBR levels. We do not claim we have the correct interpretation of the model, but only of having a consistent model which may or may not be the same as the one intended by IFLA.

[10] Consider for instance the translation of a Norwegian book into Italian, where the two contents have practically no words in common.

[11] <http://www.akomantoso.org>

[12] At least one of the authors of Renear & Wickett 2010 has discussed and criticized both the FRBR hierarchy and its relation to XML in a number of earlier publications, so perhaps their association of XML documents with FRBR Expressions is just a simplification made for the sake of their argument. However, their argument does seem to rely at least to some extent on this simplification.

[13] <http://purl.org/spar/fabio>

[14] One may be reminded of Luis Borges' short story "Pierre Menard, Author of the Quixote", where an (invented) XX century French writer re-writes word for word Cervante's Don Quixote (a XVI century Spanish writer), not by copying it, nor by imitating Cervantes, but by *becoming* a XX century French writer that through his own conceptual and philosophical growth generates the Quixote (in XVI century Spanish, of course). We guess that most reader's intuitions do not provide a clear answer whether we are dealing with one or two documents in such a case. On the Timed Abstract Object account, however, the answer is clear enough: They are two different documents.

[15] The example is taken from [Simons 1987], p. 126.

[16] [Simons 1987], p. 123. The "continuant/occurrent" distinction is (roughly) a distinction between things, substances or individuals on the one hand, and events, processes etc. on the other. Simons contends that adherents of four-dimensionalism are "cheating", i.e. they are relying on the ordinary distinction between continuants and occurrents in all their claims that the former can be reduced to the latter. He does not argue that a language where all references to continuants were consistently translated to references to four-dimensional objects and their spatio-temporal parts is impossible, but that it is hard to see how it would be comprehensible, and that no one has even begun the task of constructing such a language.

[17] It may of course be argued that this is only because such information is in fact rarely available.

[18] It may be objected that a function is not an object, but a relation. This may be true, but is hardly relevant for the argument presented here.

[19] [Swoyer 2008], p. 27. See also [Rosen 2012] and [Dorr 2008])

Claus Huitfeldt

`<claus.huitfeldt@uib.no>`

Associate Professor

Department of Philosophy, University of Bergen, Norway

Claus Huitfeldt is Associate Professor at the Department of Philosophy of the University of Bergen, Norway. He was founding Director (1990-2000) of the Wittgenstein Archives at the University of Bergen, for which he developed the text encoding system MECS as well as the editorial methods for the publication of Wittgenstein's Nachlass — The Bergen Electronic Edition (Oxford University Press, 2000).

Fabio Vitali

`<fabio@cs.unibo.it >`

Department of Computer Science, University of Bologna, Bologna, Italy

Fabio Vitali is Associate Professor in Computer Science at the University of Bologna, where he teaches Web Technologies and Human-Computer Interaction. His interests lie in models and languages for document management and hypertext support, and has published more than 60 papers in national and international venues. He is member of the W3C Working Group on XML Schema, and member of the scientific committee of several conferences and journals in Web engineering and technologies. He is author of important standards in the legislative XML Domain, and work on issues related to digital publishing, Web technologies and Semantic Web technologies.

Silvio Peroni

`<speroni@cs.unibo.it >`

Department of Computer Science, University of Bologna, Bologna, Italy

Silvio Peroni holds a degree in Computer Science at the University of Bologna. The main research interests in his current Ph.D. career include Semantic Web technologies, markup languages for complex documents, design patterns for digital documents and automatic processes of analysis and segmentation. He has published 9 scientific papers about these subjects.

Balisage Series on Markup Technologies