

When Digital Objects Change — Exactly *What* Changes?

Allen H. Renear¹, David Dubin², Karen M. Wickett³
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
501 E. Daniel Street, MC-493, Champaign, IL 61820-6211 USA
renear@uiuc.edu, ddubin@uiuc.edu, wickett2@uiuc.edu

Abstract

Formal accounts of digital objects often characterize them as *bit strings*, *graphs*, *sets*, *tuples*, *relations*, or other similar constructs from discrete mathematics. Such characterizations imply that these objects cannot undergo changes such as losing or gaining parts, or having their parts rearranged. Yet our discourse about digital objects seems, if taken literally, to imply that those objects routinely undergo such changes. One strategy for dealing with this inconsistency is to affirm an account which leaves digital objects immutable and re-locates change in the persons interacting with these objects.

Introduction

The digital information world appears to be a place of constant change. Files get larger, documents are edited, database records are modified, collections grow, files are reformatted. Such changes are fundamental to the role information plays in our lives and understanding, and learning to control and manage them is a large part of the subject matter of information science. Yet according to the standard formal accounts of digital objects these changes must be illusory. Formal characterizations of databases, records, files, XML documents, and such describe them in terms of *sets*, *strings*, *tuples*, *graphs* or other similar mathematical entities. And the changes imputed, such as addition and subtraction of parts, modification, or reorganization, are not changes that, strictly speaking, these objects can undergo. A set for instance is a collection of members, and by a fundamental axiom of set theory a particular set cannot lose or gain a member and still be the same set. Similar observations may be made about strings, relations, tuples, graphs, and the like.

We present one simple answer to this puzzle: apparent changes in digital objects are actually changes in us, in the person or persons interacting with those objects, and not changes in the objects themselves. What follows may be considered an exercise in the *conceptual foundations of information science*.

The Problem

Consider the sentence “I remember Verona.” Let it be the first sentence of the first chapter of a novel. Now suppose that the author decides to edit that sentence and revises it to read: “I remember, but dimly, Verona”. It is natural to say that the first sentence of the chapter has been changed, that it is now longer. But *what* has changed? *What* is longer? The “new” first sentence, “I remember, but dimly, Verona”, has not changed: it still consists, just as it always has, of those five words in the same order. Nor has the original sentence, “I remember Verona”, changed: it still consists of those three words, in the same order. “I remember, but dimly, Verona” is indeed a longer sentence than “I remember Verona.”, but it did not *become* a longer sentence than “I remember Verona” — it has always been a longer sentence than “I remember Verona”. It is natural to speak of sentences changing when they are edited or revised, but in fact, they do not change.

Responses

The text of the chapter changes: One response is that the sentences haven’t changed, but the larger text, the text of the first chapter, say, has changed. This will not help. The text of the first chapter is a sequence of sentences and as such simply a longer linguistic entity; so the argument we have just made applies to that text as well. We will still have our puzzle: what is it that has changed when the text of the first chapter, which began: “I remember Verona. According to my diary ...” is revised to read “I remember, but dimly, Verona. According to my diary...”? The answer cannot be either the new text of the chapter, or the original text of the chapter. And yet there are no other obvious candidates.

The computer system changes: Another response is that there is a physical object, presumably some part of the physical computer system, that is the thing that changes. It is certainly true that when a digital document is edited on a computer some portion of the information system undergoes a change. Now of course that concrete physical object cannot be *identified* with the sentence or the sentence would change whenever that physical system changed. But perhaps it is what really changes when digital objects appear to change: in that it ceases to represent one sentence and comes later to represent another. To see why this cannot be so note that there is no specific feature of the computer hardware and software that, alone, makes the physical configuration a representation of one sentence rather than another.

The “problem” is just word play: It may be argued that we have manufactured this problem by misconstruing assertions

like “The first sentence has been changed”, committing what logicians call a “scope equivocation”. Consider a coffee queue where George has just succeeded Susan as the first person in line. If we say “The first person in line has changed”, we do not mean that George has changed. True, but this is precisely our point: when a revision occurs no sentence changes, nor does any larger linguistic entity change. [Nor is a change in the physical line necessary for George to have become the first person in line: the coffee server may simply be attending to next person in order, with no physical change in the queue having occurred.]

The Solution

Our proposed solution is simple. When a sentence in a digital document is revised in the manner described no sentence changes, nor does any part of the larger text change, these all remain as they were. What is different is which text now *counts as* (Searle, 2001) the text of the novel. Prior to the revision it was the text that began “I remember Verona...”. Following the revision that text was no longer considered the text of the novel, but rather another text was considered the text of the novel, one beginning: “I remember, but dimly, Verona...”. What has changed is not the text, but which text the author considers to be the text of the novel.

Real Change

Suppose John becomes the strongest man in the world not by becoming stronger, but because Bob, who had previously been strongest, is now weaker. Apparently both John and Bob have changed. John gained the property of being the strongest man, and Bob lost that property. But the changes are not comparable. The change in Bob is a physically detectable alteration (smaller muscles) in his body. The change in John however cannot be detected by physical examination (of John), however minute. In fact it is tempting to say that John has not really changed at all, that real change requires more than gaining or losing *relational* properties, such as being stronger than someone else — it requires gaining or losing an *inherent* property, a property something has independently of its relationship to other things. Moreover, something undergoes a relational change only in virtue of some *other* thing undergoing a *non-relational* (real) change. John underwent a relational change only because Bob underwent a real change. (Mortenson, 2006)

This is more a matter of terminology rather than metaphysics. There is one sense of “change”, familiar and important, which requires more than a gain or loss of relational properties. Let’s call that sense “change_A”. To make this difference precise we begin with a definition from a widely used ontology evaluation system (Guarino & Welty, 2002, 2004):

“A property Φ is *externally dependent on* a property Ψ if, for all its instances x , necessarily some instance of Ψ must exist, which is not a part nor a constituent of x :

$\forall x L(\Phi(x) \rightarrow \exists y \Psi(y) \wedge \neg P(y, x) \wedge \neg C(y, x))$ ” (Guarino & Welty, 2000)

We can now say, more precisely, that Φ is *relational* if it is *externally dependent* upon some property Ψ . Properties that are not relational we call *inherent*. If a property Φ is inherent, whether or not some individual is an instance of Φ can be decided without taking into consideration any other properties or individuals. We understand some individual x to have undergone a change (in the most general sense) with respect to a property Φ when the truth value of $\Phi(x)$ has varied over time. We consider such a change to be a change_A when the property of interest is an inherent property.

Why Digital Objects Cannot Change

We begin with another definition provided by Guarino and Welty:

A *rigid property* is a property that is essential to *all* its instances, i.e. $\forall x \Phi(x) \rightarrow L\Phi(x)$.
(Guarino & Welty, 2000)

Guarino and Welty explain that the property of *being a student* is not rigid because it is possible for something that is a student to cease to be a student (and continue to exist). The property of *being a person* on the other hand is rigid: nothing which is a person can cease to be a person — without ceasing to be altogether.

It follows immediately from this definition of rigidity and our earlier definition of change_A that in order for something to change_A, it must have some properties that are both non-rigid and inherent. Do sentences have any such properties? Consider “I remember Verona”. Some of its properties are non-rigid, such as *being an example in a 2008 ASIST poster*. Some of its properties are inherent, such as having exactly three words, or containing the word “remember”. But none of its properties are both inherent *and* non-rigid. So it cannot change_A.

More generally: do digital objects, like digital texts, databases, records, files, and such have any inherent properties that are not rigid? Apparently not. And therefore no digital object can change_A.

Foundations for the Immutability of Digital Objects.

A better understanding of why digital objects have no non-rigid inherent properties comes from examining more closely the formal characterizations typically offered for such objects.

Strings. Digital texts are often described as strings, in the mathematical sense. But the string “abc” has the elements “a”, “b”, and “c” rigidly, and the order of elements rigidly. Editing a string cannot really be making changes in a string, but is rather substituting one string for another.

Graphs: More sophisticated accounts of digital text fare no better. Suppose a text is understood not as a string of symbols alone, but as a string of symbols in combination with a parse tree for that string. In that case the text is not a string but a graph with annotated nodes and edges. But such graphs in turn are defined in term of relations, which is to say sets of tuples. Editing a text cannot be making changes in a graph, but only substituting one graph for another.

Databases. Formally a database table is also mathematical relation. “Adding” a row to a table cannot be, literally, adding a tuple to a set, but, again, turning attention from one relation to another.

Files. A digital file is typically defined as some sort of sequence of data objects, and associated metadata. We may think of a file as a sequence of binary digits, but at other levels of abstraction it is a sequence of characters, integers, or application-specific structures. Since files can be copied, renamed, and migrated from one storage media to another, the only basis for identifying something as the same file or a different file is by the same/different object sequence characterized by the same/different metadata. And so as with strings and database relations, to “change” a file’s contents is really to indicate a different file.

Concluding Remarks.

It is common to speak as if digital objects change, and yet if digital objects are things like bit strings, sets, tuples, graphs, and such then they cannot change. We’ve presented one resolution: digital objects do not change, what changes are our attitudes, individual or collective, towards those objects. This resolution is simple, and, despite its realism, it is naturalistic, positing no entities beyond those already in the scientific worldview. It does imply that some common ways of talking cannot be taken literally, but this is hardly unusual, and there is an account of what is meant by such expressions.

Of course there are plausible alternatives for resolving the puzzle, and the authors of this paper are in fact far from certain that one of these alternatives will not provide a superior account. It may be argued for instance that digital objects are not abstract mathematical things but rather material things, or that they are some third sort of thing, or even that they are entirely fictional, idioms reifying social practice. Sources of alternative approaches include traditional nominalism and social constructivism of course, but we would also recommend recent theories of objects in the “middle-distance” (Smith 1996) or “mesoscopic world” (Smith 1998).

However whether any of these alternative perspectives provides a resolution that competes with the one presented here remains to be seen. At the moment it seems to us that these approaches are either too underdeveloped to be genuine competitors, or populate our world with unnecessarily strange new objects.

Acknowledgements

We are very grateful for comments from members of the UIUC/GSLIS Writers Group, the Electronic Publishing Research Group, and others, especially Les Gasser, Carole Palmer, Richard Urban, and Ingbert Floyd. However we have decided to proceed anyway.

References

- Guarino, N. and Welty, C. A. (2000). A formal ontology of properties. In Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management. Lecture Notes In Computer Science. Springer-Verlag.
- Guarino, N. and Welty, C. (2002). Evaluating ontological decisions with OntoClean. *Communications of the ACM*, 45(2) 61-65.
- Guarino, N. and Welty, C. (2004). An overview of OntoClean. In *Handbook on Ontologies*, Springer Verlag.
- Levinson, J. (1980). What a musical work is. *The Journal of Philosophy*, 77, 5-28.
- Mortenson, C. (2006). Change. In: The Stanford Encyclopedia of Philosophy (Downloaded December 19, 2006)
- Searle, J. R. (1995). *The Construction of Social Reality*. New York: The Free Press.
- Smith, B. (1998). The Basic Tools of Formal Ontology. *Formal Ontology in Information Systems 1998*, 19-28. IOS Press.
- Smith, B. C. (1996). *On the origin of objects*. Cambridge: MIT Press.