Beyond Digital and Physical Objects:
The Intellectual Work as a Concept of Interest for HCI

Melanie Feinberg

School of Information, The University of Texas at Austin

1616 Guadalupe St., Suite 5.202

Austin, TX 78701-1213, USA

feinberg@ischool.utexas.edu

ABSTRACT

To understand activities of personal collecting and preservation, HCI researchers have investigated why people become attached to particular objects. These studies have examined ways that people relate to physical and digital objects, observing, for example, that people tend to cherish physical objects more than digital ones. This paper proposes that the value of digital objects may inhere less in an object's identity as a particular item and more in the object's ability to provide access to an intellectual *work*. The work, a familiar concept in information studies and textual studies, designates a general product of intellectual creation that may be instantiated in many versions. (For example, Shakespeare's *Hamlet* exists in many editions and forms, which may differ in both content and carrier and yet still are all *Hamlet.*) The paper demonstrates how the concept of the work can extend research on the perceived value of digital objects. It also shows how a flexible definition of the work can reveal new aspects of a design situation.

Author Keywords

Works; texts; documents; digital media; textual studies, information studies; design; collecting; preserving; memory

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

As more traces of our intellectual and emotional activity become digital, HCI has turned its attention to the support of memory functions, with particular emphasis on personal archiving practices [12, 24, 11, 16]. To provide a conceptual base for this support, researchers have investigated the qualities that make any object significant or cherished [22, 7, 8, 10]. These studies have often noted distinctions between tangible objects and digital ones, finding that the items people select as personally valuable intertwine important memories with significant material

characteristics: the seashell from a summer vacation, the child's artwork. The physical presence of these significant objects may become even more powerful over time as the item is inscribed with the process of aging. The photograph of a beloved great aunt's vigorous youth gets brittle and fades, the favorite picture book from childhood features stains and rips from years of hard use. In contrast, digital objects seem less distinctive. After all, many digital objects can be easily replicated with little cost, whereas a seashell or crayon drawing is unique in its singular presence. And yet losing access to one's digital music collection or stash of Kindle mystery novels would certainly be traumatic. While the idea of missing a particular stream of bits seems strange, the idea of missing the expression enabled through those bits is not. I might not have an attachment to the digital file that encodes my copy of Stevie Wonder's *Ribbon in the Sky*, but if *Ribbon in the Sky* itself somehow disappeared from the world, and I could never listen to it again, I would sad indeed. Replacing my copy with another copy, though? I wouldn't think twice about it.

This paper introduces the concept of the intellectual work, a notion with a long history in information studies and textual studies, as a means of understanding objects that may exist as potentially vast sets of copies and almost-copies. The concept of the work provides a structure to organize the slew of versions that exist for many instances of creative expression. Is one's attachment to *Ribbon in the Sky* the song as written by Stevie Wonder, no matter its instantiation? Or is it to a particular performance of the song by Stevie Wonder (and not a cover version by another artist), or to a specific manifestation of the performance by Stevie Wonder (an MP3 file) or even to a certain physical copy (on the CD that my partner and I bought when we lived in Los Angeles fifteen years ago, that now has a tiny skip in it)? Does one adore plain *Ulysses* by James Joyce? Or does the edition matter? (To some readers, the Gabler edition is an abomination.) Does it have to be in print, or can it be on a Kindle? Is the real object of your affection the tattered copy you toted in high school to establish your reputation as a tortured intellectual? The concept of the work can help us identify important levels of abstraction for particular media forms (moving images, maps) and situations of use (pleasure reading, scholarly criticism). The concept of the work can also clarify the relationships that obtain between different categories of versions. I would

never say that I cherished any MP3 file. But having access to the work encoded in that MP3 file may be another matter entirely. (It may be, as well, that where I might once have described my collection of LPs as dear to me, I am now happy to be rid of those heavy boxes of vinyl records, as long as I can listen to the songs on my iPod. The different affordances of new versions may shift our allegiances.)

In this paper, I introduce the concept of the work and explain its history in the domains of information studies and textual studies, highlighting both the utility of the work for each discipline as well as difficulties caused by its inherent ambiguity. Following the track of textual studies, I use the example of documentary film footage to illustrate a flexible approach to defining the work, which I argue is key to its productive employment within HCI. I then show how using the notion of the work in this manner can extend analyses of significant and cherished objects from HCI research. I conclude by demonstrating how conceptualizing the work in different ways can reveal new aspects of a design situation, using the example of personal collections in the social media service GoodReads.

AN OVERVIEW OF THE WORK
In our daily lives, we employ the concept of the work without realizing it when we refer to the product of creative expression at various levels of abstraction. Consider the following scenario:

On my last plane flight, I read Murder at the Savoy, *a Swedish detective novel from the 1960s. The literal translation of the Swedish title is something like Police, Police, Mashed Potatoes, isn't that funny? I bought a used paperback because it was cheaper than the Kindle version.*

In the first sentence, *Murder at the Savoy,* the title, is being used in a general way, without reference to any particular version. In the second sentence, I distinguish between two different versions, one Swedish and one English, that are both still *Murder at the Savoy,* even though the Swedish version not only has a different string of words for the title, but a different meaning for the title. Here, although the sentence indicates two different sequences of symbols (in different languages) that represent the same unit of expression, *Murder at the Savoy,* there is no reference to the instantiation of those symbols in any particular format (printed text, digital text, audiobook). In the third sentence, I refer to different physical manifestations of the English version at two different levels of abstraction. One is the physical item that I bought. It is the one element in this story that is a distinct object; in this case, it is also tangible. The other is not a specific Kindle book on my or anyone else's e-reader but a broader notion of that expresses the idea of the (English) Kindle version in general.

Three levels of abstraction are commonly distinguished to bring some degree of order to this array of version types: document, text, and work [30, 28]. A document indicates a specific copy, like the paperback I read on the plane. A text

indicates different sets of symbols that are similar enough to be considered essentially the same creation. The text is abstract in that it refers to the set of symbols and not their physical embodiment, and so the same text (an English translation of *Murder at the Savoy)* can be available in different formats (paperback, Kindle). The work is the concept that links together all the documents and texts. The scholar of information organization Elaine Svenonius has described the work as that which brings together "almost the same information," and the textual studies scholar Paul Eggert has defined it as "what underwrites the sameness" between texts [27, 5]. That may sound frustratingly ambiguous. As the following sections make clear, no one has defined the work with satisfying precision. And yet the idea is profoundly intuitive and inescapably useful. If you ask me if I've read *Hamlet,* and I say Yes, you won't think I'm lying if I've read the Oxford edition while you've read the Modern Library edition. We unconsciously agree that both of these are equally *Hamlet.* In fact, it's hard to imagine talking about instances of creative expression without reliance on this most general level of abstraction. Most of the time, for most of us, just plain *Hamlet* is the best way of communicating what we care about, and if our yellowed copy was exchanged for a crisp new one, we wouldn't mind. We might not even mind if our copy disappeared, because we know many copies exist, and we can always find another. But what if someone exchanged our copy of the Oxford *Hamlet* with "No Fear Shakespeare, a modern translation"? We might say, "Hey, that's not really *Hamlet!"* We mean that No Fear Shakespeare is not as much *Hamlet* as the Oxford edition or the Modern Library edition; it's no longer the same work. And what about a Japanese translation of *Hamlet?* Our judgment then might depend on whether we speak Japanese, or perhaps on our particular views regarding literary translations.

The disciplines of information studies and textual studies both explore these distinctions. Textual scholars, whose work is associated with literary studies, are interested in what constitutes *Hamlet* to provide a basis for editing and interpretation, while information scholars (and practitioners, such as librarians and information architects) are interested in how the results of searching a document repository for *Hamlet* should be structured to facilitate a user's selection of appropriate items. The next sections distill elements from these traditions.

THE WORK IN INFORMATION STUDIES: COLLOCATION AND ARRANGEMENT OF VERSIONS FOR RETRIEVAL
The goal of assisting a patron in discriminating between editions has always been recognized in library cataloging, even though cataloging rules have traditionally focused on describing the item in hand (most typically a book). The nineteenth-century librarian Charles Cutter's objectives for a library catalog, which continue to underlie modern Anglo-American cataloging principles, assert that the catalog should "assist in the choice of edition," which it does in practice (from Cutter's day to ours) by describing edition

information, such as the publisher and publication date, in the catalog record [4].

Over time, however, and especially as collections became larger, it became apparent that bringing together multiple editions in the catalog was a tricky problem. Again since the time of Cutter, catalogs have prioritized three access points: author, title, and subject (in the days of card catalogs, each of these access points had its own set of cards, with items filed in the three ways). While one might initially think that the title would provide sufficient collocation, different editions are often given different titles. Think again of *Hamlet:* one might well have *The Tragedy of Hamlet, Prince of Denmark; Hamlet; Shakespeare's Hamlet;* and so forth among many possible variant titles. Too, *Hamlet* might be contained within another work: *Selected Plays, Shakespeare's Tragedies,* and so on. And there may be other items with the title *Hamlet* that aren't written by Shakespeare at all. By the 1960s, the cataloger Seymour Lubetzky, one of the architects of the Anglo-American Cataloging Rules (the foundation of modern cataloging guidelines) explicitly described the collocation of editions of a particular work as a primary goal of the catalog [15]. Martha Yee expands that a catalog should enable library users to identify documents that are both effectively equivalent or perhaps even preferable to what users were originally searching for: someone looking for the third edition of a textbook may be pleasantly surprised to see that a newer fourth edition is available, for example [32]. Moreover, items that are not the textbook being sought, but that may have similar titles, should be excluded from the search results. Through its abilities to describe and group documents as part of a single work, the library catalog (or, indeed, any document retrieval system) should actively assist users in mapping the appropriate document space and refining information needs.

The current structure of the library catalog (all libraries in the U.S. use the same conceptual structure, the same file format, and the same guidelines for creating records) does not accommodate works incredibly well; each record is still for a particular physical item. Today, collocation of editions occurs through the use of authority files that provide controlled vocabularies for subject terms, author names, and, occasionally, what are called uniform titles (these are most commonly used to relate translated titles with the ones in the original language). One tries to locate all the editions of *Hamlet* by searching the author file for William Shakespeare, which retrieves all the records associated with that author, and then searching those records for *Hamlet* in the title (at my university library, there is no uniform title associated with *Hamlet*). The recall of such a search are invariably incomplete, and options for grouping and sorting the results (over 500 at my university library) are limited.

An entity-relationship model to clarify work-related levels of abstraction

To address such issues, in 1998 the International Federation of Library Associations (IFLA) introduced an entity-relationship model for catalogs based on the concept of the work [9]. This model, Functional Requirements for Bibliographic Records (FRBR) proposes a set of four bibliographic entities of increasing concreteness: works, expressions, manifestations, and items. Each entity has associated attributes and potential identified relationships. In the FRBR model, expression is similar to the more commonly used "text," while manifestation refers to the output of a particular printing (a print run, but also an issue of CDs, e-book, or whatever media). The FRBR model is incorporated into the latest cataloging rules, Resource Description and Access (RDA), recently completed and due to begin adoption by the U.S. Library of Congress (LC) in March, 2013. Most U.S. libraries will follow LC's lead, because most American libraries obtain LC's cataloging records and adapt them for local use. Currently, however, the benefits of FRBR can only be approximated in most library catalogs. The OCLC WorldCat system, a union catalog that aggregates the holdings of 72,000 libraries worldwide, enables some grouping of editions by exploiting properties of the current record structure, but this service is incomplete and inconsistent [21].

Austlit, the online Australian Literature Resource, provides an illustrative example of how the FRBR model would facilitate document retrieval. Austlit was built around the FRBR model and adopts the work, and not the item, as the record unit. A collaborative venture of the National Library of Australia and Australian universities, each bibliographic record in Austlit brings together all the expressions (called versions in Austlit) of a work, as well as all the manifestations (called publications in Austlit) of each expression. (Austlit does not contain item records; it links to item records held by participating libraries.) A sample record for Patrick White's novel *Voss* brings together information about 22 expressions (mostly different translations) represented in 43 manifestations (all printed books but issued by different publishers); it also includes references to related works such as an opera based on the novel and several excerpts contained in literary anthologies [1]. An excerpt from this sample record appears in Figure 1. With a catalog based on works, such as Austlit, it is easier for the information seeker to understand relationships between versions and to discriminate between them.

What kind of entity is a work? Or is it a relationship?
While the FRBR initiative has cemented the importance of the concept of the work for document description and retrieval, it has not provided a rigorous definition of the work, nor has it established a clear set of principles for determining when a particular expression is part of an existing work or when it should be considered a new work. FRBR describes a work as "a distinct intellectual or artistic creation" and as an entity, although an entity without any

Figure 1: Excerpt of Austlit record for a work.
Each numbered item is an expression; some expressions have
several manifestations.

physical or even symbolic representation. Most discussions of the work within information studies characterize it as an "abstract entity," which is congruent with traditional ideas from textual studies (summarized in the following section). As an example, Smiraglia claims that the work is the set of ideas that "lies behind" all the work's expressions [26]. All of the 22 expressions of White's novel *Voss* from Austlit share a common group of ideas, although the words in each expression might be totally different, and these shared ideas are the work. Both the appeal and difficulty of this sort of definition is that it conceives of the work as a static ideal; it implies that although different expressions might come into being, the work that these expressions represent doesn't change. In this vein, Svenonius remarks on the "abstract, Platonic" nature of the work as a concept [27].

Some scholars, however, conceptualize the work as a relationship or category instead of an entity. O'Neill and Vizine-Goetz define the work as "a set of related texts with a common origin and content," and Renear and Dubin assert that the FRBR entity is logically a relationship [23, 27]. These less rigid notions of the work are echoed by more recent scholarship in textual studies (also summarized in the following section).

Difficulties in specifying what constitutes a work lead to downstream confusion in determining which expressions constitute new works. FRBR guidelines, which do not provide systematic rationale, can appear arbitrary. For example, a sound recording of *Hamlet* is an expression of the original work (and thus as equally *Hamlet* as the printed

Modern Library edition), but the filmed version of a live performance of the play is a new work. Problems associated with not being able to articulate what constitutes a work are particularly apparent for non-book materials, such as maps, motion pictures, and scientific models [18, 31, 3].

Digital materials can be especially complex, because significant differences in underlying code may be undetectable in the user experience, and the sheer number and scope of versions can be dizzying. To demonstrate this, McDonough, et al enumerate the set of versions associated with a simple, early computer game (Adventure); they are unable to adequately differentiate these versions with the current set of FRBR entities [17]. In conducting their case study, McDonough and colleagues argue that the distinctions they seek to reveal are necessary information for the software studies scholar, and so they should also be pertinent for the information professions, who support such research. But McDonough, et al equally acknowledge that the typical game player is not concerned with this level of detail in characterizing the work. Within information studies, pragmatic goals for uniform description of information-bearing objects have supported the idea of single, consistent notion of the work, with a standard set of accompanying entities. But one can alternately view the work from a more flexible perspective, adapting the characteristics that define and relate versions to accommodate certain forms (such as software) and use situations (such as scholarly research as opposed to personal entertainment). The arc of scholarship within textual studies, as described in the next section, supports such an approach. In subsequent sections, I argue that a flexible take on the concept of the work can enrich the notions of significant and cherished objects that have been a significant element of personal archiving research in HCI. A standard, uniform approach to defining the work, as realized in the FRBR model, is appealing its relative fixity. However, I suggest that a flexible concept of the work, where the idea of what counts as an essentially equivalent version may vary according to situational factors such as context of use, is ultimately more useful for HCI.

THE WORK IN TEXTUAL STUDIES: ESTABLISHING THE BASIS FOR INTERPRETATION

Similarly to information studies, textual scholarship, a branch of literary studies, aims to describe the universe of versions of a work and the relationships between versions. However, the traditional goal of textual scholarship has been to discriminate between more and less authoritative versions of literary works, and to assemble evidence that allows the construction of more reliable versions (critical editions). The foundations of textual scholarship lie in the historical transmission of manuscripts from scribe to scribe across centuries of copying, where changes across copies are inevitable (and, with extended passage of time, are potentially gigantic). Medieval scholars, for example, use specialized knowledge of alphabets and scripts (paleography, the study of handwriting) to trace the

sequence of versions over time and place, with the aim of assembling a version as close to the original as possible. Modern textual scholarship in English is epitomized by the case of Shakespeare's plays. Early printed versions were notoriously full of printers' mistakes, cuts, and so on. The textual critic's job is to assemble the chain of versions (of which none is actually "correct") and provide various forms of evidence through which an authoritative version of (say) *Hamlet* can be created. Textual scholars see their efforts as crucial to literary interpretation. Without authoritative versions, the conclusions of literary criticism are suspect.

Traditional textual studies: authorial intentions as the primary regulating principle that shapes a work
G. Thomas Tanselle articulates the traditional concept of the work in textual studies as the true expression of the author's intentions [28]. A central component of this view of the work is that any chain of textual transmission (presumably, even from the brain of the author to the initial manuscript) is rife with errors. The intended text—the work—may never have actually taken shape in a particular document, either in manuscript or in print. Nonetheless, it is the textual editor's duty to construct a text that best approximates the ideal of the work, using the existing textual variants and associated historical evidence to do so. The variant texts and the documents that contain them may reflect the work to some degree and provide clues to the ultimate nature of the work, but the work exists somehow independently of them. (The prevalent concept in information studies of the work as an abstract entity is quite similar.) Accordingly, Tanselle can assert that there is no distinction between "literary" texts and other texts—that, for example, it is absurd to have one edition of William James for philosophers and a different edition of William James for literary scholars. The work is the same for all readers, and all serious readers should prefer a text that makes the best possible claim for reflecting the work [28].

Postmodern textual studies: the work as a dynamic, collaborative process
As literary criticism has come to appreciate the socially and historically constructed elements of all interpretive activities, as suggested by the writings of theorists such as Michel Foucault, Roland Barthes, and many others, the association of meaning with authorial intention has weakened. With these ideas permeating literary studies, textual scholars have reexamined their function, along with ideas of the work. Jerome McGann has persuasively contested the association of textual transmission with errors and inevitable corruption, proposing a sense of the work as a collaborative process between authors, editors, publishers, and others. The changes requested by an editor are, in McGann's view, just that, changes, and not mistakes that take a particular text further from the ideal work. For McGann, the work is dynamic and emergent: "a series of specific 'texts,' a series of specific acts of production, and the entire process which both of these series constitute" [19, p. 52]. Current textual scholarship has married these ideas

with new technologies to produce critical editions as digital collections. The aggregation and encoding of textual variations, along with interfaces to structure comparison and interpretation, constitutes a strand of digital humanities research. Examples of such collections include McGann's Rossetti Archive, Clement's collection of the modernist poet Baroness Elsa von Freytag-Loringhoven, and the collaborative Walt Whitman Archive edited by Folsom and Price [20, 2, 6]. The forensic approach to literary analysis originated by Matt Kirschenbaum is also an extension of textual studies; the McDonough, et al case study on code-level differences between versions of Adventure is a collaboration between digital humanists/textual scholars and information studies researchers [13, 17].

In a complementary response to this general reevaluation of the notion of authorship, Eggert introduces contextual and historical elements to the idea of work-as-process, using the restoration of Leonardo da Vinci's Last Supper as one example [5]. Previous restoration attempts from other eras are now deemed inappropriate by the current restorers; they obscure the artist's original vision. In contrast, the current restoration, which uses modern technology to isolate the original fragments, enables, in the view of the restorers, the original vision to be again revealed. In the current context, from the perspective of the restorers, the older restorations are no longer part of the work, but the recent restorations are. And yet some critics contend that the new restorations reveal not Leonardo's vision but a contemporary notion of what the original should look like, and the restoration method itself is conditioned by current ubiquity of image close-ups and detail views.

The need for a flexible definition of the work
The definition of the work, from this example, appears both dynamic and debatable. Selecting a particular definition is in itself an act of interpretation. This suggests that any definition of the work for a particular context must be explicated and defended. If there is no ideal essence motivating the idea of the work, then the structure of the work must be explicitly determined for particular cases. The work becomes a generic category, almost a placeholder; it must be fleshed out before it can be used productively. Moreover, authorship becomes only one potential regulating principle for defining a work and determining its boundaries. For some forms of expression, and some use situations, additional principles may be more salient in determining the set of versions that makes up a particular work. In the next section, I use the example of documentary film footage to illustrate how the concept of the work might be specifically and productively defined for a certain context. I then show how this flexible approach to the work can extend ideas of significant objects as conceptualized in HCI, and how it can also be used to inform design decisions for systems that enable the collection and display of digital materials.

CASE STUDY FOR CONTEXTUAL DEFINITION OF THE WORK: DOCUMENTARY FILM FOOTAGE

In my view, the concept of the work is best defined as a category, or a relationship between particulars. The regulating principles of the category, or the properties that structure and organize the category members, may vary in different situations. In contemplating potential instantiations of the work as category, it is helpful to acknowledge that the most common and recognizable example of a work remains that of a written text created by a single author with a specific initial publication date, like Raymond Chandler's *The Big Sleep* or Agatha Christie's *Murder on the Orient Express*. The combination of author and title forms the nexus of the regulating principle that the work represents. While different editions might make small changes from one to the other (for example, some printings of *The Big Sleep* use the spelling "okey" for "okay"), the substance of each text is consistent across versions, presumably as written by the author. In other situations, however, the variation between versions can be much greater, such as in subsequent versions of a textbook, and authorship may change as well. *Gray's Anatomy* is on its fortieth edition, and it is no longer supervised by Dr. Gray, who died many years ago. In such a case, very little may remain the same from the first version to the most recent. The continuous publication history is perhaps more important to the character of this work as a related set of versions than either the author or the title (which was initially *Anatomy of the Human Body*, not *Gray's Anatomy)*.

Action, not authorship, as the regulating principle for a work of documentary film footage

Authorship is perhaps even less indicative of what usefully circumscribes a set of "almost the same" materials when considering certain combinations of both media and situation. An illustrative example comprises documentary film or video footage of a particular event, such as an apartment fire, tennis match, or birthday party. While the edited presentation of such an event, such as ESPN's coverage of a particular Wimbledon match, would exhibit a form of institutional authorship and as such have a distinct identity from, say, the BBC's coverage of the same match, the raw output of all the various cameras recording the proceedings (including spectators' cell phones, and so on) can be seen, from a certain perspective, as a single work.

In this case, it is the event itself, or the action, that forms the regulating principle around which the set of versions that instantiates the work revolves. The centrality of action as a cornerstone of identity for moving images has been suggested for fictional narratives as well: Andrea Leigh, in discussing episodes of television series, in particular *I Love Lucy,* suggests that the principle of action is of primary importance for entertainment-oriented information seeking, because nobody remembers the episode titles and writers; everyone instead thinks of "the one where Lucy works at the chocolate factory," and so on [14]. But such claims are even stronger for unedited recorded images of events like "the gas explosion on Payne Avenue from last November" or "Gayatri's first birthday party at Dolores Park" where the role of authorship tends to be minimized even if the person (or people) who recorded the footage is known. Let's say that three party guests sent digital video of the birthday girl blowing on her candles to Gayatri's parents. The relationship between the clips, as formed by their shared subject, the event of the birthday party, seems much stronger than the difference between them as being filmed by different people. Moreover, while the family may well want to retain video evidence of the event, they might not have any special connection to one of the clips as opposed to another, if they all included the same activities. The attachment is to the work, here defined as video of the candle-blowing, and not to a particular expression or item.

Use context and the definition of a work

To demonstrate further why a flexible approach is necessary in determining the most appropriate regulating principle to structure the versions of a work, let's add another level of complexity to the birthday video scenario. What if one of the serendipitous videographers at the birthday party happened to be a famous film director, say Martin Scorsese? Wouldn't the idea of authorship be salient in that case, and wouldn't there be a particular attachment to his clip? Indeed, such potential variability is why I propose that the use situation must contribute to any particular definition of the work as well. It may initially seem strange to adapt the determination of what kinds of text are properly *Hamlet* according to a use context. Shouldn't the idea of what is or is not *Hamlet* be something that we can all rely upon? Certainly, while McDonough, et al, note that the needs of the specialist researcher may demand more levels of discrimination between versions of the game Adventure than those required by the general player, their goal is to propose modifications to the FRBR model, in order to make it a more flexible overarching system, and not to propose flexible models for different situations [17]. Still, observations of the dynamic, historical nature of the work as a concept, as noted by McGann and Eggert, hint that no single model will ever be flexible enough [19, 5]. Of course No Fear Shakespeare will never be properly *Hamlet* for a literature scholar, and yet it might well be so for a nervous high-school student—at least, according to some principles of pedagogy. In the case of the birthday party video, if Scorsese were Gayatri's Uncle Marty, and if he was one of several habitual video contributors, there would be more of a reason to treat his clip as but another almost interchangeable version of Birthday Girl Blowing Her First Candle. If not, then there would be a reasonable argument for considering Scorsese's video as another work entirely. Similarly, if the use context was not the personal archive of family memories but digital assets for a stock video company, then the camera angle and lighting might contribute to the appropriate regulating principle for the work, and the level of detail at which the

action was described would be less important (baby's first birthday and not Gayatri's birthday).

With such potential for variation, one might begin to wonder about the ultimate utility of the work as a concept at all. Is it too nebulous to provide a rigorous and systematic analytical lens upon the diverse range of expressive artifacts in the world? Although I certainly acknowledge the difficulties in its application, I nonetheless contend that the idea of the work can help us understand the potentially vast expanse of artifact versions in revealing and thoughtful ways. Despite its inherent ambiguities and the need to think carefully about what it means for any particular combination of media and use context, and to think carefully about the level of precision appropriate for differentiating between versions, the concept of the work is a compelling means of understanding document ecologies, particularly in the digital realm, where ease of copying makes for extensive sets of similar files. The work provides a structure through which we can clarify the nature of relationships between documents and the associated meaning that those documents may hold for their creators, their keepers, and their seekers. In the next section, I demonstrate how the idea of the work as developed through this paper can further the analysis of current HCI research on the preservation of significant and cherished objects.

WHEN IS THE OBJECT OF AFFECTION NOT ACTUALLY AN OBJECT? THE WORK IN THE CONTEXT OF HCI RESEARCH ON KEEPING AND CHERISHING

In Kirk and Sellen's excellent depiction of the values enacted through home-based practices of collecting and keeping, they introduce a typology of objects that their participants cherish: physical, digital, or hybrid [12]. Hybrid items are storage media through which both digital and analog content can be accessed (videotapes, music CDs, LPs). While these distinctions are valuable to recognize, the notion of the work and its accompanying layers of abstraction (expressions, manifestations, and items, to use the FRBR terminology, which is easier to apply to non-text material) can add an additional level of complexity to this characterization. Kirk and Sellen describe the hybrid in particular as a case in which the expression or experience enabled through the object is what matters to the owner, and not the physical presence of the item itself, noting that ". . . the actual VHS or tape cassettes used for storage held no sentimental value whatsoever, but the content was considered to be very precious" [12, p. 1014]. It is the song on the CD (a la *Ribbon in the Sky)* that is more often cherished, and not the CD itself. But works with potentially numerous versions appear in the physical and digital categories as well. The physical category includes items such as newspaper or magazine cuttings, books, and printed photographs, while the digital category includes such items as video clips, e-mail messages, and digital photographs. Across the typology, for items that may find expression in multiple versions, at what level of abstraction does attachment lie? Is it to the work in general

(say, Yotam Ottlenghi's recipe for sweet corn polenta), to a particular expression or set of expressions (from the U.S. cookbook *Plenty* or from the UK Guardian newspaper column, or any UK version or any U.S. version), to a particular manifestation or set of manifestations (the U.S. Kindle cookbook, or any U.S. version that can be printed), or to one unique item (torn out of the UK Guardian and kept in the kitchen recipe file)? Answering such questions can sharpen our understanding of what it means to cherish anything; the focus of our attachment may reside in a class of similar objects, instead of a particular object. Where Kirk and Sellen assert that "we can never equate a digital copy of a physical object, no matter how veridical, with its original," this may depend on the level of abstraction at which value is located. Value may inhere in an item; or it may be connected to a work, with many potentially interchangeable items to choose from. If the area of concern in preserving an item is really to maintain convenient access to a work, then attachment to any particular copy is diminished when this access in maintained via some other means, particularly when output in different formats is possible. When I moved to a new house last year, for example, I got rid of my collection of music CDs; it was no longer necessary to store those items, because I had access to my preferred expressions of those songs on my computer and iPod, and the new devices can also be connected to the stereo receiver. My attachment to the works expressed through the songs never changed, but I no longer had need of the CD copies to maintain my access to the works.

The work of Golsteijn and colleagues also provides another demonstration of this [8]. Golsteijn et al used Kirk and Sellen's typology to analyze results from a similar study (although where Kirk and Sellen used home tours to gather data regarding significant objects, Golsteijn et al conducted focus groups where participants brought photos of selected objects). Golsteijn et al note particularly cases where the distinction between physical and digital seems to no longer matter. They describe the example of siblings scanning a set of printed family photographs so that each could have a complete set; ultimately, both siblings retained the digital images instead of the printed versions. On the one hand, it remains worthwhile to use such examples to explore the complementary affordances of different media, as Golsteijn, et al do. But we can also use such cases to examine the potentially different values that accrue to levels of abstraction: work, expression, and so on.

It is may also be possible to use these ideas to obtain a more precise sense of the value people identify in mass-produced designed items that undergo regular versioning, as with consumer electronics and computers. It may be reasonable to think about the next version of an iPhone in the same way that one would think about the next version of a textbook or a travel guide. In both cases, the newest versions are typically more desirable—in other words, the regulating principle of the work is focused around recency. Am I happy to replace my local restaurant guide with this

year's updated version? Everything else being equal, of course I am (say if someone gave me the new version, and I didn't have to pay for it). When my university replaces my MacBook with a newer model, do I complain about wanting to keep my older, slower one? Not when the IT staff handles all of the migration tasks for me. Would I always be happy with such exchanges, and what would lead me to refuse them? Through being aware of the concept of the work and its accompanying levels of abstraction, HCI can more readily formulate such questions and investigate their answers.

As another example, these and similar studies ask participants to share significant objects with the researchers. And yet, as has been discussed, a digital photo may be valued for the expression it presents, not for its status as an object. The terminology of objects may discourage participants from selecting items that are valued for the access to expression that they enable more than their material qualities. Petrelli and Whittaker, for example, also directed participants, in home tours, to select special things in physical and digital form [24]. Petrelli and Whittaker were surprised that few people selected photos as significant physical objects. Only 16 percent did so, and Petrelli and Whittaker observed that selected photographs were often "unique or irreplaceable"; that is, not available in multiple versions. In the digital realm, Petrelli and Whittaker's participants selected items that they or others had created or that documented personal experiences (that is, they tended to select family photographs and videos, their own e-mail exchanges, and digital content created by children, and not their iTunes libraries or funny videos of cats collected from around the Web). In both the physical and digital realms, the focus on "special things" may have encouraged participants to concentrate on item-based characteristics instead of higher levels of abstraction. A cat video, in other words, may be special, but not as a thing, and a music collection may be significant, but not identified as such on a home computer if it is also available on a iPod and the laptop in the office.
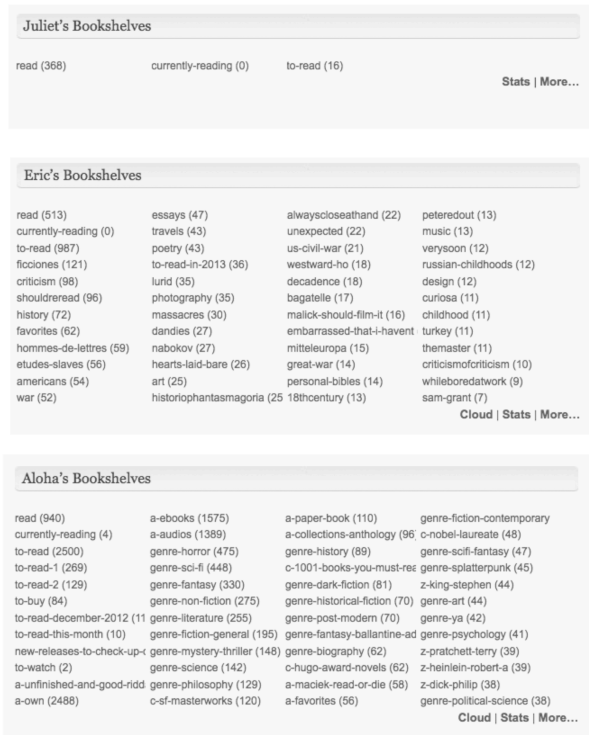
Consideration of work-related levels of abstraction can also complement studies of personal information management in office and academic settings, such as Whittaker and Hirschberg and Kaye, et al [29, 11]. These studies, too, focus attention on the distinctions between physical (often paper, in this case) and digital items. Whittaker and Hirshberg, for example, identify factors that lead people to retain paper files of documents even when moving offices. While a good portion of retained paper documents were unique items, and another segment of retained documents were in the "to-read" pile, 36 percent of retained paper documents were available elsewhere. The most prevalent forms of rationale for keeping these materials were to preserve convenient, reliable, and persistent access (a variation involved distrust of other access methods). Here, the value seems to be at the level of the work, although one's own print versions are perceived to enable access

better than other versions. But it's the access that's important, not the paper nor the personal storage. Another factor for retention, reminding, relies on the affordances of paper as a medium (a manifestation-level attribute): having a paper document laying around on one's desk keeps the actions associated with it in mind, and browsing articles in files reminds one why they are important. A final factor, sentiment, is associated with item-level characteristics: materials used in one's dissertation, for example. There are memories bound up in the saved items, even though participants recognize that there is nothing about the item's appearance that would make such associations apparent to someone else. Access to the work isn't actually important here, it's the sentiment bound in the materials themselves. While Whittaker and Hirshberg's analysis is perceptive and detailed, the discussion of retention factors could gain additional nuance by considering which level of abstraction is important for each factor, and why this is so. Sometimes the affordances of a particular medium, such as paper, are important, but when the value is at a higher level of abstraction, then alternate media and associated access mechanisms may be appreciated, not just tolerated. However, when the significance lies in the item (even if the traces of attachment are not to be observed in the material itself), then no substitution will be acceptable. Seen from this perspective, the behavior of the administrator with the "almost paperless" office described in Kaye, et al's study of personal document collections is quite reasonable [11]. This administrator kept paper versions only of those materials that he had in some way participated in creating. These items were valued as unique objects; the digital library of documents on the administrator's computer, on the other hand, was valued as a means of accessing works, not because of any item-level significance.

IDENTIFYING PARTICULAR CONCEPTIONS OF THE WORK TO ILLUMINATE A DESIGN SITUATION

In this section, I briefly illustrate how the concept of the work can be used to comprehend a design situation and inform the development of alternative feature sets. As an illustrative example, I consider the domain of collecting, which encompasses a range of activities from the keeping of cherished objects as repositories of memory to the maintenance of personal files for information access. Here, I focus on the collection and display of metadata records as enabled through the social media service GoodReads. In the GoodReads environment, users assemble collections of book records into personal catalogs. The use of books is less central to this example than the idea of a collection of surrogates: analogous systems include Pinterest (collections to linked images) and even something like Facebook (collections of friend profiles). In all these cases, the collections that users create can form the basis for commentary and conversation with the community. The

Juliet's Bookshelves

read (368)   currently-reading (0)   to-read (16)

Stats | More...

Eric's Bookshelves

read (513)	essays (47)	alwayscloseathand (22)	peteredout (13)
currently-reading (0)	travels (43)	unexpected (22)	music (13)
to-read (2500)	poetry (43)	us-civil-war (21)	verysoon (12)
ficcciones (121)	to-read-in-2013 (36)	westward-ho (18)	russian-childhoods (12)
criticism (98)	lurid (35)	decadence (18)	design (12)
shouldreread (96)	photography (35)	bagatelle (17)	curiosa (11)
history (72)	massacres (30)	malick-should-film-it (16)	childhood (11)
favorites (62)	dandies (27)	embarrassed-that-i-havent	turkey (11)
hommes-de-lettres (59)	nabokov (27)	mitteleuropa (15)	themaster (11)
etudes-slaves (56)	hearts-laid-bare (26)	great-war (14)	criticismofcriticism (10)
americans (54)	art (25)	personal-bibles (14)	whileboredatwork (9)
war (52)	historiophantasmagoria (25	18thcentury (13)	sam-grant (7)

Cloud | Stats | More...

Aloha's Bookshelves

read (940)	a-ebooks (1575)	a-paper-book (110)	genre-fiction-contemporary
currently-reading (4)	a-audios (1389)	a-collections-anthology (96	c-nobel-laureate (48)
to-read (2500)	genre-horror (475)	genre-history (89)	genre-scifi-fantasy (47)
to-read-1 (269)	genre-sci-fi (448)	c-1001-books-you-must-rea	genre-splatterpunk (45)
to-read-2 (129)	genre-fantasy (330)	genre-dark-fiction (81)	z-king-stephen (44)
to-buy (84)	genre-non-fiction (275)	genre-historical-fiction (70)	genre-art (44)
to-read-december-2012 (11	genre-literature (255)	genre-post-modern (70)	genre-ya (42)
to-read-this-month (10)	genre-fiction-general (195)	genre-fantasy-ballantine-ad	genre-psychology (41)
new-releases-to-check-up-c	genre-mystery-thriller (148)	genre-biography (62)	z-pratchett-terry (39)
to-watch (2)	genre-science (142)	c-hugo-award-novels (62)	z-heinlein-robert-a (39)
a-unfinished-and-good-ridd	genre-philosophy (129)	a-maciek-read-or-die (58)	z-dick-philip (38)
a-own (2488)	c-sf-masterworks (120)	a-favorites (56)	genre-political-science (38)

Cloud | Stats | More...

Figure 2: Different use contexts for GoodReads shelves imply different ideas of the work for a collection of book metadata.

collection-building features of GoodReads, however, make it a good example for this scenario.

One of GoodReads' features is the *shelf,* essentially a category for grouping collection items (here, typically records for books owned by the collection creator, usually described at the level of a FRBR manifestation). There are three default shelves: r*ead, currently-reading,* and *to-read.* Users can create their own shelves and identify these custom shelves with labels (or tags). Differences in shelf use suggest that users' collections constitute different types of works.

Some GoodReads users have significant collections but do not create additional shelves. The user identified as "Juliet Evans," with a collection of 384 books, only includes the GoodReads default shelves. Other users, however, create extensive shelf systems. For example, even though the GoodReads system does not facilitate the creation of multiple category levels, the user "Aloha" created a multidimensional shelf structure using shelf naming conventions as grouping devices. Other users, such as "Eric," devise evocative, idiosyncratic shelf categories such as "bagatelle" and "lurid." Figure 2 shows the shelf systems for the users identified as Juliet Evans, Aloha, and Eric. (Note that the figure can only show a small selection of Aloha's total number of shelves, which extend to the hundreds, including a shelf for every author.)

I propose that in this example, there are two ideas of the work, as distinguished by two different contexts of use. For GoodReads users who do not create additional shelves, or who create very few shelves, the regulating principle of the work centers on the set of records describing each entered book. However, for the users with extensive shelf systems, the regulating principle of the work is the category structure as applied to the records. In terms of design implications, a new version, or expression, of the GoodReads catalog for users like Juliet would occur only when an item is added or deleted. For users like Aloha and Eric, however, a new version would involve a category change: addition, deletion, or assignment (a new record might be created in the context of a new assignment, or an assignment might involve existing records). Each of these ideas of the work might lead to different modes for saving, displaying, and exporting a user's catalog. Currently, GoodReads does not support any catalog versioning or archiving. Especially for users like Aloha and Eric, for whom the value of the catalog entwines deeply with the shelf system, this would seem to represent a useful design opportunity, as revealed through a work-oriented analysis.

CONCLUSION

This paper demonstrates how the concept of the work, as drawn from information studies and textual studies, can extend HCI research on the perceived value of digital objects, and how attending to the various levels of abstraction that structure a set of versions can provide another layer of understanding regarding the relationships between physical and digital objects. While the work, with its inherent ambiguity, requires flexibility in its definition and application for particular situations of use and forms of media, it can nonetheless provide a keen analytical lens to facilitate our conceptual grasp of collecting activities and attitudes.

REFERENCES

1. Austlit, the Australian Literature Resource. Available at http://www.austlit.edu.au/. Last accessed August 30, 2012.

2. T. Clement, ed. In transition: selected poems by the Baroness Elsa von Freytag-Loringhoven. Available at: http://digital.lib.umd.edu/transition/index.jsp. Last accessed September 2, 2012.

3. A. Coleman. Scientific models as works. In *Works as Entities for Information Retrieval.* R. Smiraglia, ed. New York: Haworth Information Press,129–160, 2002.

4. C. Cutter. *Rules for a dictionary catalog.* 4th ed. Washington: Government Printing Office, 1904.

5. P. Eggert. Where are we now with authorship and the work? *Yearbook of English Studies* 88–96, 1999.

6. E. Folsom and K. Price. Walt Whitman Archive. Available at: http://www.whitmanarchive.org/. Last accessed September 2, 2012.

7. S. Gegenbauer and E. Huang. Inspiring the design of longer-lived electronics through an understanding of

personal attachment. *Proceedings of ACM DIS 2012,* 635–644, 2012.

8. C. Golsteijn, E. van den Hoven, D. Frohlich, and A. Sellen. Toward a more cherishable digital object. *Proceedings of ACM DIS 2012,* 655–644, 2012.

9. International Federation of Library Associations (IFLA) Study Group on the Functional Requirements for Bibliographic Records. *Final Report.* Munich: K.G. Saur, 1998. Available at http://www.ifla.org/publications/functional-requirements-for-bibliographic-records. Last accessed August 28, 2012.

10. H. Jung., et al. How deep is your love: deep narratives of ensoulment and heirloom status. *International Journal of Design* 5(1) 59–71, 2011.

11. J. Kaye, et al. To have and to hold: Exploring the personal archive. *Proceedings ACM CHI 2006,* 275–284, 2006.

12. D. S. Kirk and A. Sellen. On human remains: Values and practice in the home archiving of cherished objects. *ACM Transactions on Computer-Human Interaction* 17(3): 1–43, 2010.

13. M. Kirschenbaum. *Mechanisms: New media and the forensic imagination.* Cambridge, MA: MIT Press, 2008.

14. A. Leigh. Lucy is "enceinte": the power of an action in defining a work. In *Works as Entities for Information Retrieval.* R. Smiraglia, ed. New York: Haworth Information Press, 99–127, 2002.

15. S. Lubetzky. *Principles of cataloging.* Report for U.S. Office of Education grant OE-1-7-071089-4284. Los Angeles,: UCLA Institute of Library Research, 1969.

16. C. Marshall. Digital copies and a distributed notion of reference in personal archives. In *Digital Media: Technological and Social Challenges of the Interactive World.* M. Winget and B. Aspray, eds. Lanham, MD: Scarecrow Press, 2011.

17. J. McDonough, et al. Twisty little passages almost all alike: applying the FRBR model to a classic computer game. *Digital Humanities Quarterly* 4(2), 2010. Available at http://digitalhumanities.org/dhq/vol/4/2/000089/000089.html. Last accessed August 28, 2012.

18. S. McEathron. Cartographic materials as works. In *Works as Entities for Information Retrieval.* R. Smiraglia, ed. New York: Haworth Information Press,181–191, 2002.

19. J. McGann. *A critique of modern textual criticism.* Chicago: University of Chicago Press, 1983.

20. J. McGann, ed. Rossetti Archive. Available at: http://www.rossettiarchive.org/. Last accessed September 2, 2012.

21. OCLC. WorldCat catalog. Available at http://www.oclc.org/worldcat/default.htm. Last accessed August 31, 2012.

22. W. Odom, et al. Understanding why we preserve some things and discard others in the context of interaction design. *Proceedings of ACM CHI 2009,* 1053–1062, 2009.

23. E. O'Neill. and D. Vizine-Goetz. Bibliographic relationships: implications for the structure of the catalog. In *The Conceptual Foundations of Descriptive Cataloging.* E. Svenonius, ed. San Diego, CA: Academic Press, Inc., 167–195, 1989.

24. D. Petrelli and S. Whittaker. Family memories in the home: contrasting physical and digital mementos. *Personal Ubiquitous Computing* 14(2) 153–169, 2010.

25. A. Renear and D. Dubin. Three of the FRBR Group 1 entities are roles not types. *Proceedings of ASIST 2007.*

26. R. Smiraglia. *The nature of "a work": implications for the organization of knowledge.* Lanham, MD: Scarecrow Press, 2001.

27. E. Svenonius. *The intellectual foundation of information organization.* Cambridge, MA: MIT Press, 2000.

28. G. T. Tanselle. *Textual criticism and scholarly editing.* Charlottesville, VA: University of Virginia Press., 1990.

29. S. Whittaker and J. Hirschberg. The character, value, and management of personal paper archives. *ACM Transactions on Computer-Human Interaction* 8(2): 150–170, 2001.

30. P. Wilson. The second objective. In *The Conceptual Foundations of Descriptive Cataloging.* E. Svenonius, ed. San Diego, CA: Academic Press, Inc., 5–16, 1989.

31. M. Yee. The concept of work for moving image materials. *Cataloging and Classification Quarterly* 18(2): 33–40, 1993.

32. M. Yee. What is a work? Part 4: cataloging theorists and a definition abstract. *Cataloging and Classification Quarterly* 20(2): 3–24, 1999.