What Exactly is an *Item* in the Digital World?

Ingbert R. Floyd

Graduate School of Library and Information Science, University of Illinois at Urbana Champaign, 501 East Daniel Street, MC-493, Champaign, IL 61820-6211 USA; Tel: (217) 721-3171; ifloyd2@uiuc.edu (please correspond with both authors)

Allen H. Renear

Graduate School of Library and Information Science, University of Illinois at Urbana Champaign, 501 East Daniel Street, MC-493, Champaign, IL 61820-6211 USA; Tel: (217) 265-5216; renear@uiuc.edu (please correspond with both authors)

IFLA's Functional Requirements for Bibliographic Records (FRBR) is a model of the bibliographic universe. Although initially its application to the digital world appears to be straightforward, upon closer examination puzzles arise. One is that within the digital world it is surprisingly difficult to say exactly what FRBR items really are. On the one hand, the ontological candidates for items (concrete physical states of the computing system) are rarely identified and treated as items in practice — even though they may indeed be affirmed as items in theoretical discussions. On the other hand, objects that manifestly fail to meet the basic ontological criteria for FRBR items are commonly treated as if they *are* items. We describe this situation and, based on a re-factoring of FRBR into a set of roles (relationships) rather than a set of entity types, explore two possible resolutions. One, favored by the second author, is consistent with the ontology implicit in the original FRBR vision, but allows assignment of item attributes and roles to things that are not items; the other, favored by the first author, is a radical departure from the underlying FRBR ontology, but preserves the original attribute assignments and roles.

FRBR

The Functional Requirements for Bibliographic Records (FRBR) is a "conceptual model of the bibliographic universe" (IFLA 1998). The FRBR Group 1 entity types are *work*, *expression*, *manifestation*, and *item*. *Works* are characterized as "a distinct intellectual or artistic creation", *expressions* as "the intellectual or artistic realization of a work in the form of alphanumeric, musical, or choreographic notation..." (a particular text for instance), *manifestations* as "the physical embodiment of an expression" (an edition for instance), and *items* as "a single exemplar of a manifestation". Appropriate attributes are assigned to each entity, for instance a *work* might have a genre, an *expression* a language, a *manifestation* a typeface, and an *item* a condition. Works, expressions, and manifestations are abstract objects, items are concrete physical things.

The Concept of an *Item* in the Digital World[1]

Conceptualizations of the bibliographic universe vary in the number of bibliographic kinds they recognize as entities (Svenonius 2000; IFLA 1998), but in every case they include an entity that corresponds to the FRBR *item*. That is, an entity which is a concrete physical object that in some sense records or carries information. Upon reflection, the importance of items is obvious: for one thing, one of the main functions of a traditional library is to collect, house, maintain, and distribute items. And while information itself may be abstract, as long as we are material beings in a material world, it would seem that that without physical items there can be no recording, communication, or reception, of information. It is not surprising that the analysis of the nature of the item figures prominently in the classic discussions of foundational issues in LIS (e.g., Buckland 1997; Buckland 1998).

So what in digital libraries (or repositories, archives, etc.) corresponds to the FRBR *item*? There is no new puzzle about what are *works* in the digital world, as these are the very same conceptual things that are works

in the wider analog world. The identification of *expressions* and *manifestations* is perhaps somewhat more challenging (Renear et al 2003), but does not seem to profoundly challenge the spirit of the FRBR framework. The case of *item* however is more difficult. Some of these issues were raised by Buckland (1998), but the specific problem below has so far received little attention.

Items as Patterned Matter and Energy in the Computer System

In addition to the definition given above FRBR characterizes an *item* as follows:

> The entity defined as *item* is a concrete entity. It is in many instances a single, physical object (e.g., a copy of a one-volume monograph, a single audio cassette, etc.). (IFLA 1998 23)

In most scenarios of information use in the digital world there is no visible, discrete, concrete physical object comparable to a copy of a book. Nevertheless, we can infer the existence of an entity that seems an appropriate candidate for classification as an item: individual physical states of the relevant portions of the computing system. We will refer to these physical states as instances of *patterned matter and energy* (PME)[2]. However the identification of *items* as PMEs is challenged by the fact that in the digital world individual PMEs are rarely assigned any of the important attributes (such as provenance, identifier, or condition) and relationships that FRBR items supposedly can have (IFLA 1998, 48-50, 63-82).

It may not be apparent at first why we do not speak of, for instance, the *provenance* of a PME. After all, we speak of digital objects as being transferred from one place to another, one server to another, etc. But it is not usually the case that a particular concrete PME is being moved from one place to another (as it would be if we shipped a compact disk through the US Mail). Rather, a PME that exists in one place is used to create a copy (which is *another* PME) elsewhere. This suggests why PMEs, although they are natural candidates for items in virtue of their status as concrete physical objects and their role in physically instantiating information, are not usually identified as first class bibliographic objects in digital contexts. Not only are PMEs perceptually distant and, to most users, obscure, but also individuating distinct PMEs, whose periods of existence are often very short and difficult to ascertain, seems irrelevant to bibliographic control in the digital world (though not to other concerns of digital librarians, such as, most obviously, material preservation). Similarly we believe that a careful analysis of other common item-level attributes, such as item identifier, condition, etc., will also reveal that these are rarely applied, strictly speaking, to a PME, but rather to some related abstract object, perhaps corresponding manifestation or expression.

Items as Files

If PMEs are not treated as items in the digital world, then what is? What has a provenance, an item identifier, was copied, was transferred from one place to another? The answer often appears to be a *file*. "Where did this file come from?" asks a digital librarian in a forensic mood, and continues: "please make a copy of it" or "move it to the H-Drive" and so on. Files, it seems, are item-like. But what is a *file*?

The answer to this question is not obvious. There are patterns of practice and discourse around the word "file" but it is far from clear whether there is a single univocal concept behind them, much less what that concept is. Of course, within particular contexts (operating systems, data structure theory, etc.) there are exact definitions of "file", but those are of little help here: we wish to know what a file is in the sense in which the term is actually used by digital librarians and users. Several possibilities suggest themselves. The first is that a file is an abstract object, a sequence of bits or characters — which is clear enough, but would make the rhetoric of creation, destruction, location and such metaphorical and blur the distinction between item and manifestation. Another is that files are logical fictions such as the "the average plumber", which raises the question of what entities will be involved in their explication. A third is that a file is a "free standing" social object, like a debt or a corporation (Smith 1993), but the nature of such social objects is controversial.

Interlude: Refactoring FRBR Types as Roles

Following Renear and Dubin (2007) we see expressions, manifestations, and items not as fundamental types of entities, but as roles that certain other entities take on under certain circumstances of linguistic convention and social intention, much as described in Searle (1995). So, in the right social circumstances, *symbol sequences* are expressions, *physical kinds* are manifestations, and *physical objects* are items.[3] This revision of FRBR solves some puzzles of entity assignment (Renear et al 2003), better conforms to current best practice in ontology design (Guarino & Welty 2000), and provides a useful framework for comparing two quite different resolutions of the problem of the item in the digital world.

Solution A: Items in the digital realm really are PMEs

This solution maintains that a fundamental aspect of being an item is being a concrete physical thing that provides the material substrate for the instantiation of manifestations, expressions, and works. Therefore items in the digital world must be PMEs just as FRBR claims. More specifically, items are PMEs in the role of exemplifying a *manifestation* that embodies an *expression* that realizes a *work*. It is true that, in the digital world, items will in many respects not be as prominent, nor function in exactly the same way, as items as envisioned by FRBR. In particular, they may have only a few of the assigned attributes or relations. According to this solution, however, those attributes and relations are not essential to the role of *being an item*.

What about the fact that FRBR item attributes or similar properties are routinely used to characterize *other* things, such as, files? Solution A holds that files in this sense are, for the most part, logical fictions (like the average plumber), and the locutions in which they figure are, for the most part, metaphors. There is no entity whatsoever which was sent when a “file” is “sent”, no entity that is moved when a “file” is “moved”, and no entity which is “compressed” when a “file” is “compressed”. Such locutions are convenient, even necessary for efficient conversation about things digital, but underneath these fictions, metaphors, and polysemy, is a real structure of material things, general abstractions, and socially determined roles as described in the refactored version of FRBR

Solution B: Items in the digital realm are files

This solution holds that items are the things, whatever their nature (physical, abstract, or metaphorical), which play the role in bibliographic control that FRBR assigns to items. In this account, the relevant item-specific attributes and relations are what *define* the role of being an item. FRBR itself provides such an account:

> Defining *item* as an entity enables us to separately identify individual copies of a *manifestation*, and to describe those characteristics that are unique to that particular copy and that pertain to transactions such as circulation, etc. involving that copy.
>
> Defining the entity called *item* also enables us to draw relationships between individual copies of *manifestations*. (IFLA 1998, 23)

This approach holds that the fundamental nature of any FRBR entity is the role it plays in bibliographic control, not its ontological status. And in the case of most commonly used computing systems, the thing that fills the role of the FRBR item is the “file”. If we insist on calling PMEs items, then we will need to develop a new role (or entity) that is neither *item* nor *manifestation* and which has the attributes currently assigned to items; this new bibliographic construct will correspond to the common conception of a file. Otherwise the model will not reflect practice.

What does this solution say about the fact that the concept of *file* seems poorly defined? It declines to make a formal definition of a “file”. It is sympathetic to the “file as social object” approach, but rejects this view as too deterministic—clearly, a programmer may have a different definition of “file” than a novice computer user. However, since the term “file” with its multiple interpretations allows communication between people who hold different definitions, it is a classic boundary object (Star & Griesemer 1989). The exact ontological nature of a

boundary object may be mysterious, but the benefit is that they can be functionally used as a black box—i.e., assumed to be an entity (or an entity playing a role). This approach reflects Wittgenstein's view that words do not have definitions, they only have uses: the meanings of the terms in our bibliographic and information processing vocabularies are fully constituted by how we use those terms in our every-day speech, thoughts, and actions.

Conclusions

The initial source of this puzzle is that the FRBR account of the *item* entity combines two families of features that appear to be necessarily coextensive, but that in fact vary independently, at least in the digital world: items are physical objects that carry information *and* items are the things which have certain attributes and relationships relative to bibliographic control. In order for FRBR to be effectively applied in the digital world, these two families of features need to be distinguished and reanalyzed. We have described two approaches to this. Of course the specific strategy chosen to support system design and standards integration will be a question of pragmatics, intuition, and social agreement; our purpose here is to point out the conceptual implications of the different approaches.

Acknowledgments

This paper owes a special debt to the insights of Dave Dubin and to discussions (from 2003-2007) at the GSLIS Metadata Roundtable (led by Carole Palmer and Michael Twidale. It also draws liberally on the discussions conducted in joint meetings of the Electronic Publishing Research Group (lead by Dubin) and Topics in Knowledge Representation seminar (lead by Renear) in the spring 2006 (participants included Yunseon Choi, Jin Ha Lee, Karen Medina, Sara Schmidt, Richard Urban, Xin Xiang, and Oksana Zavalina). Finally, an earlier draft was considerably improved by presentations to the GSLIS Research Writing Group (led by Dave Dubin), where it received useful comments from Thomas Dousa and others members of GSLIS.

References

Buckland, M. K. (1997). What is a Document? *Journal of the American Society for Information Science, 48*(9), 804-809. Retrieved from http://www.ischool.berkeley.edu/~buckland/whatdoc.html on 5/27/2007.

Buckland, M. K. (1998). What is a Digital Document? *Document Numerique (Paris), 2*(2), 221-230. Retrieved from http://www.ischool.berkeley.edu/~buckland/digdoc.html on 5/27/2007.

Cooper, A. (1995). The Myth of Metaphor. *Visual Basic Programmer's Journal, 5*(6). Retrieved from: http://www.cooper.com/articles/art_myth_of_metaphor.htm on January 24, 2006.

Searle, J. R. (1995) *The Construction of Social Reality*. New York: The Free Press. 1995

Guarino, N., & Welty, C. A. (2000) A formal ontology of properties. In *Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management*, R. Dieng and O. Corby, Eds. *Lecture Notes In Computer Science, vol. 1937*. (Springer-Verlag, London) 97–112.

IFLA Study Group on the Functional Requirements for Bibliographic Records (1998). *Functional Requirements for Bibliographic Records: Final Report (FRBR)*. Munchen: K. G. Saur. Retrieved from: http://www.ifla.org/VII/s13/frbr/frbr.pdf on January 23, 2006.

Mohnkern, K. (1997). Visual Interaction Design: Beyond the Interface Metaphor. *SIGCHI, 29*(2). Retrieved from: http://acm.org/sigchi/bulletin/1997.2/vid.html on January 24, 2006.

Renear, A. (2006). Is an XML Document a FRBR Manifestation or a FRBR Expression?—Both, Because FRBR Entities are not Types, but Roles. *Extreme Markup Languages 2006*, Montreal, Quebec.

Renear, A., Dubin, D. (2007). Three of the Four FRBR Group 1 Entity Types are Roles not Types. *ASIS&T,*

Proceedings of the Annual Conference.

Smith, B. (1993). From Speech Acts to Social Reality. In *John Searle*, Ed B Smith, Cambridge: Cambridge University Press.

Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*(3), 387-420.

Svenonius, E. (2000). *The Intellectual Foundation of Information Organization*. Cambridge, MA: MIT Press.

---

[1] We use “digital world” as it is commonly used, to indicated the practices, tools, and policies associated with modern electronic digital computing, ignoring for the moment the more accurate distinctions that might be made between pairs such as digital/analog, physical/abstract, electronic/mechanical, real/virtual, etc.

[2] The phrase “patterned matter and energy”, deftly coined by Dave Dubin, refers to any arbitrary concrete physical phenomenon, without being specific as to what kind of phenomenon it is (matter, energy, some combination of the two, or some particular material event). The phrase emphasizes, by using the adjective “patterned” rather than the noun “patterns of…” that the entity referred to is, whatever else it may be, an individual concrete instance and not a repeatable abstraction.

[3] For instance, in the right social circumstances, some combination of ink and paper is an item, or part of one; in other circumstances it is an accident.