WHAT IS A DIGITAL ASSET?

Richard Pearce-Moses
Persistence of Memory: Stewardship of Digital Assets
Boston : 1 November 2005
Tucson: 5 December 2006

ABSTRACT

"Digital asset" can be used to refer to a wide range of formats and genres, some of which have close analogs in the real world and others that exist only in virtual space. Pearce-Moses will explore the meaning of "digital asset" in terms of functions and characteristics and how different contexts give different senses to the concept. A richer understanding of the use and components of digital assets can help frame the larger question of stewardship.

Over the past several years, I've been lost in the world of words, writing the Society of American Archivists' new glossary. In the process, I've become something of an amateur lexicographer, analyzing the way people use words to discern what those words mean. That experience serves as the backdrop for this presentation. As I take up the question, 'What is a digital asset?,' I'm simultaneously thinking about how we use language and – in particular – how we use jargon.

[SLIDE 2]
THE MEANING OF MEANING

Language is a powerful thing. Used carelessly, words confuse. Used wisely, they inform. As such, it's important to use them carefully, to use them thoughtfully. That means paying attention to what we're saying. That's what I hope to do today: to pay attention to what we mean by the phrase 'digital asset'.

Choosing the right word means fully understanding its meaning, and that is not a simple task. Generally, we don't learn definitions from a standard dictionary. We learn definitions from context, and that context varies for each of us. As a result, our sense of what a word means is ad hoc rather than formal. . . . The more we focus on the formal details of meaning, the more elusive it becomes. If I use the word blue, many will think of the color, while others will think of the mood. Given the many shades and tones of the color, it's possible that none of those thinking of the color will be thinking of the exact same color. The general rule is that when defining a term, precision and consensus are at odds. At a high level, people may agree on meaning. But when they get to the details, there's a lot of disagreement. We may never agree on a precise, specific definition of 'digital asset', even if we share a consensual understanding of the phrase.

In addition to the definition denoted in a dictionary, a word's connotation adds nuance and mood. The right word will ring true in context – or sound silly if misused. A number of years ago, Bruce Handy wrote about the use – or abuse – of the word 'postmodern' in odd contexts. For example, why would editors at *Elle* describe a ski parka as postmodern? After contacting several writers who had used the term, he concluded that postmodern had become "culturespeak, short for Stuff That's Cool in 1988. It's the current

version of groovy – except that using it makes you sound smart."[1] The phrase 'digital asset' has different connotations from a number of similar terms I'll be talking about.

[SLIDE 3 – TRIGGER EACH LINE SEPARATELY]
JARGON AND CANT

As Handy observed, people use language to impress as much as anything. That's particularly true in marketing, where it's important to distinguish one product from something else. To motivate people to buy, ads pitch their products as something new or different. To do that, they often abandon established language for something new. Information technology seems to be especially subject to such trendy cant. 'Data processing' begat 'information processing', 'information processing' begat 'knowledge management', and 'knowledge management' begat 'wisdom management'. I suspect soon we'll see products for 'omniscience management'.

[SLIDE 4]
GENERAL DESCRIPTION TABLE

The public relations firm Hodskins, Searls, and Simone developed a Generic Description Table that allows anyone to easily construct a new phrase that sounds impressive but is vague, at best.

Adverbs	*Adjectives*	*Adnouns*	*Nouns*
Alternatively	Advanced	Access	Alternative
Automatically	Architected	Automated	Analysis
Backwardly	Bridging	Breakthrough	Business
Digitally	Dedicated	Design	Document
Virtually	Parallel	Paradigm	Protocol

Table 1. Entries from the Generic Description Table.[2]

[SLIDE 5]
YET ANOTHER GDT

I use this table to help set the stage for the challenge of defining 'digital asset', and I offer my own table based on a review of literature on the web.

Adjectives	*Nouns*
Content	Asset
Data	Object
Digital	Resource
Information	
Knowledge	
Learning	
Wisdom	

Table 2. Generic Online Thing Table

1 "The Rise and Fall of a Great American Buzzword," *Spy* (April 1988).

2 John A. Barry, *Technobabble* (MIT Press, 1991), p. 195-201.

Based on my research, there is little difference between a digital asset, a digital object, and a digital resource. The terms are often used interchangeably. In some works authors distinguish the terms, but different authors distinguish them differently.

Digital asset was frequently used in the context of the monetary value of the information, when there was an emphasis on the investment of time, thought, and funds in the creation, management, and maintenance of the digital information. But digital resource was often used in with a similar connotation.
[SLIDE 6]
EXAMPLES OF DIGITAL ASSETS

We use the phrase digital asset – and its variants – to include a wide range of things. Lexicographers are descriptive, not prescriptive. They derive definitions from the way words are used, not from a theoretical ideal. The result, of course, is that there is no neat, logical taxonomy, and definitions can be a bit messy.

[SLIDE 7 – CLICK THROUGH GAG LINE TWICE]
WHAT IS A DIGITAL ASSET? – WHAT EVER YOU WANT!

At the risk of sounding flip, I will answer the question, "What is a digital asset?" with the observation that it can mean whatever you want it to mean. Especially if you're in marketing and trying to make your product sound like the next, best thing. We can define a digital asset (or a digital object or a digital resource) a number of ways.

In a technical context, reflecting the connotations of object-oriented programming, it is "A unit of information that includes properties (attributes or characteristics of the object) and may also include methods (means of performing operations on the object)."

In a more general context, it can be defined simply as "An item as stored in a digital library, consisting of data, metadata, and an identifier."

[Slide 8]
WHAT IS A DIGITAL ASSET?

I'd like to propose the definition on the screen as a starting point . . .

> *A unit of information that includes substantive content and metadata that documents the context and structure of the information and that provides for the administration, discovery (access and use), and preservation of the information.*
> *~ The unit is based on artificial boundaries*
> *~ Connotes value based on time and money invested in creation, management, and maintenance.*

A definition is a useful place to start, but a definition is also a very rarified abstraction. It's one thing to know that a diamond is crystallized carbon. It's another thing to consider the many facets of a diamond. I'd like to spend some time looking at the characteristics and functions that make up the facets of a digital asset.

[SLIDE 9]
FUNCTIONS (FOR USE)

How we understand anything depends on context. How we use something provides a central context for understanding the object and can help us identify some of its characteristics.

Curation

Curatorial functions are those activities we perform when building and managing a collection of digital assets. You'll see a list of some of those functions on the screen.

Selection and acquisition: When looking at the universe of things we might acquire, we may restrict our working definition of digital assets to something that we can manage. If we can't easily bring in an entire database, we might not consider the entire database as an asset. However, we might acquire information in the database as discrete units, treating those units as separate assets.

Rights management: If we want to make money from our assets, we may want to use rights management schemes to protect our investment. However, if we are making our collections freely available, we might not even think of a number of characteristics of digital assets such as the many layers of intellectual property, watermarks or banners.

Discovery

For many of us, the phrase 'digital asset' immediately brings to mind online collections. Libraries, archives, and museums want to get their collections on the web, and the digital surrogates are digital assets.

As with any collection, we must provide some mechanism to help people find materials they're interested in. We do that, what we now call discovery, with – another word with a new meaning – metadata, or as Robert Gorman has called it, "cataloging for men".

Patrons will use these assets for information and for data analysis. In the first case, they parse the content by reading it, thinking about it, and adapting it for their own purpose; unstructured text and images are very useful here. But people looking for data to analyze and mine are looking for datasets. If the dataset is in the form of scanned text, it's going to be easy to read but hard to analyze. If it's in an HTML table, users have some chance of transforming the data into a format that can be analyzed. However, if the digital asset is a delimited flat file, it may be hard to read, but easy to analyze.

Preservation

Preservation is a part of curation, but of such importance in the digital environment that it merits separate treatment.

Many digital assets are reproductions of born-analog materials. Digitization not only provides the possibility of increased access, it protects the original from handling. At the same time, there's reasonable debate as to whether digitization is appropriate for preservation reformatting. However, that question is moot when it comes to born-digital materials. Unless we want to convert those items to paper or film, we must keep the bits alive over time.

The challenge of long-term digital preservation identifies a number of characteristics that at first seem trivial because they are so common, but which may have a significant impact on the content – and

the value – of the digital assets. Changes in the methods to encode and decode the digital information into human-readable form can affect the appearance and functionality of the assets.

[SLIDE 10]
CHARACTERISTICS

Independent of function, we can distill many more characteristics of a digital asset inherent in its material – or shall we say virtual – nature.

Keep in the back of your mind that digital assets that serve as surrogates of material things have two sets of characteristics. Of the asset itself, and of the thing it stands for. For example, the asset may be a tiff or jpg, but it may represent a letter, a photograph, or a map. The asset may have been created by a technician using an Epson scanner, but the original may have been created by George Eastman using a Kodak.

One of my beliefs is that – at an abstract level – there's nothing new under the digital sun. There's something analogous in the material world to everything in the virtual world. For example, contemporary PKI used to sign and authenticate documents is conceptually similar to chirographs, a technique from the Middle Ages used for the same purpose. This continuity and similarity is a good thing because it means we can transfer a lot of our existing knowledge into cyberspace. We'll have to translate that knowledge, but we have a good starting point.

To the extent that a digital asset is, in essence, a virtual document, we can start with our understanding of the characteristics of documents – at an abstract level – to think about digital assets. One of the basic things we know about documents is that they have content, context, and structure.

[SLIDE 11]
CONTEXT

Context means the asset (or the thing it represents) has a place in a larger world. It may be part of a collection and may have order within that collection. It may have been created for a specific function. The asset doesn't exist in isolation.

Provenance
Original order
Production (place, date)
Acquisition
Authenticity
Rights and permissions
Confidentiality and privacy

[SLIDE 12]
STRUCTURE

A digital asset often has an enormously complex structure. At a minimum it is a bitstream, a sequence of zeros and ones that is meaningful only to a machine.

Extent: The asset has size and boundaries

If digital assets are to be accessed through the web, size matters. Where's the value in transmitting a very large, high resolution digital image when few computer monitors can display the image at anything better than 96 dots per inch?

Data formats

Those zeros and ones may be stored in a number of different formats. A digital image could be a TIFF, a JPG, or a GIF. A text document could be flat ASCII; marked up text, such as HTML or XML; or proprietary, binary formats, such as Word or WordPerfect.

Storage media and formats

The bitstreams in those data formats must ultimately be stored on some physical media, whether spinning disks, tape, or something else. Each of those physical media may encode the digital assets differently.

[SLIDE 13]

COMPONENTS

At the most atomic level, a digital asset is nothing more than a sequence of zeros and ones. The essence of a digital asset, though, is the structure of those zeros and ones, how they are organized and interpreted so that they are meaningful.

Digital assets may combine text, images, and sound. The assets is made up of multiple assets. Often the object is bounded in terms of more familiar, if virtual, parameters: pages for a book, dimensions for an image, and running time for a sound or moving image recording.

Which raises the question: is the asset encapsulated or distributed? Are the separate assets stored together as a single asset, or does some magical software combine assets stored separately? One of the values of PDF is that it can encapsulate many formats so that we only have one asset to manage. XML can be used for the same purpose.

Encapsulating components in a single asset allows those components to be fixed, a critical quality for a record. If the components are drawn from other sources, it's possible that those components will change over time – often intentionally. The information may be constantly updated from real world sources or maybe with intentional randomness, with the results that the asset may change each time you look at it.

The quality of being variable or fixed is a bit different from the qualities of an asset being dynamic or static; for example, it may be an animated or an unmoving gif.

[Slide 14]

IDENTIFIER: THE ASSET HAS A NAME.

We need to be able to distinguish one object from another using a short handle or name. Librarians and archivists tend to think of this in terms of a title. It's important to give even untitled works some sort of identifier so that we can represent the asset in a list or citation.

For assets that are surrogates, we may use an accession number as an identifier, linking the digital surrogate with its original.

While humans may be comfortable with any number of works called “Untitled,” machines want a distinct title to distinguish one asset from all others. One common technique is to use a universally unique identifier (UUID), or as known in Microsoft implementations, a globally unique identifier (GUID).

[Slide 15]

CREATOR AND ROLE: THE ASSET HAS SOME ORIGIN.

All artifacts have some sort of creator. A book has an author, a symphony a composer, a photograph a photographer. In some instances, there may be several people sharing roles; a song has a lyricist and a composer, an anthology has several authors. Even objects produced by an automatic process are the product of the people who made the machine.

[SLIDE 16]

CONTENT (SUBJECT)

In many ways, the most important quality of a digital asset is its content. Content is the information that is the heart of the digital asset.

If a museum digitizes a rock, the image is a representation of the subject -- the rock. It may be nothing more than a specimen, a diamond. Or it may be a very specific example, such as the Hope diamond.

However, the content of synthetic works often has two very distinct levels. Consider a photograph of a horse by Alfred Stieglitz. On a representational level, it is ***of*** a horse. However, when one learns that the horse represented is clearly a gelding (something that had to be pointed out to this city boy) and one reads that the title of the work is ‘Spiritual America’, the work takes on a discursive, metaphorical meaning, a subject the photograph is ***about.***.

We can talk about content in terms of common concepts – people, groups, places, activities, and dates. But we should also consider the content that we add to the digital assets. Our descriptions add significant value as a record of what we know about the asset.

[Slide 17]

FORM AND GENRE (NON-TECHNICAL)

We often seek content based on categories. Those categories fall into two categories: physical and intellectual.

Physical: Text, still and moving images, sounds. Raw data.

Intellectual: Fiction, encyclopedias, reports, blogs, correspondence, fantasy, photojournalism.

[SLIDE 18]

PRESCRIPTIVE DEFINITION

Given that the phrase digital asset is used with such a wide range of meanings, a descriptive definition is not terribly useful. So let me end with something of a prescriptive definition that incorporates many of the characteristics I’ve touched on.

Digital information

- valuable for invested time and money
- used for a variety of purposes, within allowable rights and permissions
- with content (substantive data and metadata)
- with context (provenance, original order, administrative metadata)
- with structure (intellectual and data formats, preservation metadata)

Observation

Finally, I'd like to conclude with an observation. Speaking as an archivist, one of the challenges of the digital era is the emphasis on digital assets as items. Archivists emphasize aggregates – collections and record groups, series and subseries, and folders. Archivists typically manage and describe at the item level only if the materials are of exceptional importance.

In many ways, this approach is tactical rather than principled. We will never have the resources to manage and describe each individual item in our holdings. With the explosion of digital assets, I question if an emphasis on item-level management is realistic. At the Arizona State Library and Archives, we came to realize that traditional bibliographic models would not work for materials published on state agency websites; there were just too many materials out there.