UC Curation Center

Digital Curation Foundations

Rev. 2.0 (*draft*) – 2015-02-28

Abstract

Digital curation is a complex of actors, policies, practices, and technologies that enables meaningful consumer engagement with authentic content of interest across space and time. To ensure that it is using its curation resources in the most productive manner, the University of California Curation Center (UC3) has modeled the curation domain to provide a consistent, comprehensive, yet parsimonious conceptual foundation for the planning, implementation, and evaluation of its manifold activities. The UC3 Sept model builds upon, and attempts to consolidate, prior efforts such as Kahn and Wilensky, FRBR, OAIS, NAA performance model, PLM, PREMIS, BRM, ICO, SPOT, and NDSA levels of preservation. It also draws upon relevant concepts from cognitive psychology, information science, and semiotic theory. The model considers curated content with respect to five distinct semiotic dimensions: semantics, syntactics, empirics, pragmatics, and dynamics, which refer respectively to content's underlying abstract meaning or affect, symbolic encoding structures, physical representations, behaviors, and evolution through time. Correspondingly, there is a hierarchical typology of accumulating content utility: entities, artifacts, articles, commodities, assets, and heirlooms, which are respectively existential, intentional, purposeful, meaningful, useful, and reliable digital objects. Content engagement is modeled in terms of productive, managerial, and consumptive roles and loci of concerns co-existing within a continuum of originating, organizing, and pluralizing dimensions, which respectively encompass the establishment of, imposition of structure upon, and extension of the reach and consequence of curated content. Curation strategies are modeled in terms of six high-level imperatives: predilect, collect, protect, introspect, project, and connect. The Sept model components and terminology can be used to make precise yet concise statements regarding curation intentions, activities, and results.

Categories and subject descriptors *Applied computing ~ Digital libraries and archives • Applied computing ~ Enterprise modeling • Information systems ~ Information lifecycle management • Information systems ~ Data management systems*

Keywords *digital curation, digital preservation, domain model, content model, object model, semiotics, lifecycle, continuum, planning, strategy*

1 Introduction

Digital curation is a complex of actors, policies, practices, and technologies that enables meaningful consumer engagement with authentic content of interest across space and time.

A given unit of digital content is of *interest* if it can be readily distinguished from the larger universe of potential alternative content on the basis of consumer criteria and *authentic* if it is what it purports to be. A consumer's engagement is *meaningful* if the content can be successfully exploited for some desired purpose, whether creative, consumptive, or additive, at a time and place and in a manner of the consumer's choosing.[1] It is possible that this purpose may be fulfilled only at some considerable spatio-temporal distance from the point of the content's creation; however, regardless of how distant that point, the consumer's purpose is not necessarily constrained to conform to the original intention of the content's creator. Rather, every engagement is uniquely situated with respect to the *subjective* context of the content's curatorial framing and the consumer's experience, expertise, and expectation.

The curation attributes of enablement, interest, authenticity, and meaningfulness are a contemporary restatement of traditional stewardship concerns as exemplified by Ranganathan's "laws" of library science (Ranganathan, 1931). The first law, "Books are for use," shorn of its biblio-centricity, is fundamentally concerned with *utility*, that is, the use for purpose underlying successful engagement with a meaning-bearing artifact. The second and third laws, "Every reader his book" and "Every book its reader," are fundamentally concerned with ensuring an effective *connection* between content and consumer as a consequence of an expression of consumer interest. The question of whether the "book" is what it purports to be is one of authenticity, a traditional concern of archival diplomatics that is especially important in the digital realm given digital content's ease of mutability (Ross, 2007). Mutability of a different sort is also implicated in Ranganathan's fifth law, "The library is a growing organism," which is fundamentally concerned with *change*, corresponding to curation concerns with content's extension across space and time. The fourth law, "Save the time of the user," is fundamentally concerned with *service* and corresponds to the imperative of curating agents providing their customers with tools and services that effectively meet their intellectual, behavioral, and technical expectations. Underlying all of these concerns is the notion that curation encompasses both the preservation *and* use of digital content, which are complementary rather than disparate activities: preservation ensuring use *over* time while use is dependent upon preservation *up until a point* in time (Rusbridge, 2008).

Curation outcomes naturally lie along a continuum of possible results largely dependent upon the degree to which appropriate resources are available. Some of the factors pertinent to resource allocation decisions are intrinsic to the content itself, such as size, format, structure, and presence (or absence) of self-describing metadata; others are extrinsic, such as collection development policies, curatorial assessments of value, degree of uniqueness or ubiquity, ease of reacquisition or regeneration, availability of appropriate standards, best practices, and tools, staffing levels, and competing demands on finite organizational resources. Given the inevitability of resource constraints it is important that curating institutions make fully informed decisions to allocate (or withhold) resources and perform (or forgo) activities. This will enable institutions to plan and implement effective solutions that maximize curation utility, that is, provide the highest *overall* level of curation outcomes for the largest body of content with the least effort, while simultaneously expending *proportionate* effort towards any *given* unit or class

[1] The term "engagement" is used deliberately to encompass the widest possible range of use cases: interactive as well as passive, dynamic as well as static, creative as well as consumption, immersive as well as presentational, gestalt as well as granular, and inter- as well as intra-relational.

of content based on its assessed value and institutional priority.

2 Modeling the domain

Curation decisions should be made with respect to an underlying theory or conceptual domain model. A domain model is an abstraction of fundamental descriptive and explanatory principles sophisticated enough to explicate past events and conditions and predict the consequences and efficacy of future decisions and actions (Reynolds, 1971; MacKenzie et al., 2006). It is useful to build up such a model incrementally from first principles in order to ensure comprehensive scope, self-consistency, and parsimony.[2] The University of California Curation Center (UC3) has created a domain model that relies on a number of useful antecedents for model construction drawn from digital curation and preservation theory and practice as well as cognitive psychology, information science, game theory, and semiotic theory. All models, however, are at best *idealized* representations of *nominal* domain concepts. The simplifying assumptions and abstractions inherent to any modeling effort often may be somewhat incommensurate with pertinent real world detail and any actual curation entity or condition may not fully conform to explicit or implicit model categories or definitions. Nevertheless, the model provides UC3 with a useful conceptual map, analytical framework, and descriptive vocabulary applicable to its multifarious curation activities.

2.1 Perception and cognition

Since the ultimate goal of digital curation is to "deliver" content, that is, some coherent unit of intellectual meaning or emotional affect, to a human consumer, any persuasive model of the curation domain must be cognizant of innate human psychological and cognitive faculties. This human focus is important since even in cases of intermediating technical systems the ultimate agency in any consumptive act always resides in a controlling human actor (Dallas, 2007). A more precise way of describing content consumption is that the (analog) sensory *perception* of the (digital) content ultimately invokes a *mental reformulation* of the content's underlying meaning or affect by the consumer (Flouris and Meghini, 2007). Broadly speaking, the process by which meaning or affect is acquired progresses through several discrete stages (see Figure 1):[3]

1. An abstract unit of *content* is …
2. Realized by physical *stimuli*, which are …
3. Perceived by one or more *sense* modalities, …
4. Apprehended as an uninterpreted sense-image or *percept*, …

---

[2] The principle of parsimony, most famously articulated as Occam's Razor, posits that the minimally sufficient explanation is to be preferred over more complex alternatives (Baker, 2010). By definition simpler than other approaches, a parsimonious model should be inherently more amenable to both intuitive and empirical validation. It's important to note, however, that *simple* is not synonymous with *simplistic*; an appropriate conceptual economy can nevertheless be universal in applicability and rich in expressive power.

[3] In this and subsequent figures, named solid shapes represent atomic *entities*, labeled arcs represent directional *relationships* between entities, and dashed lines enclose *aggregate* entities.

5. Interpreted in the specific *context* of the consumer, and ultimately …
6. Experienced as *meaning* or *affect*.

In defining the final crucial transition from perception to cognition it is important to recognize that content consumption is an inherently *semiotic* act.

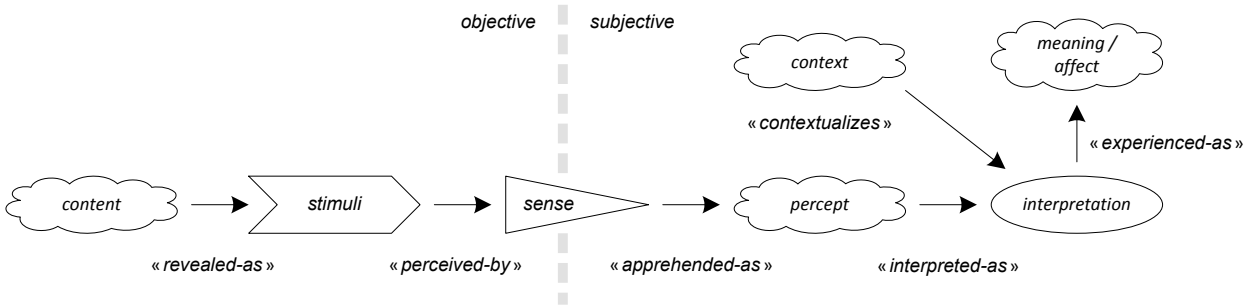

Figure 1 – Human cognition model

2.2 Curation semiotics

Semiotics is the science of sign systems and signification, that is, things that *carry* meaning or affect and the ways in which that meaning or affect are *represented* and *communicated* (Eco, 1974; Nöth, 1990). In Peirce's triadic theory of semiotics a *sign* is something that "stands in" for something else, for someone, in some manner (Peirce, 1932). More formally, a Peircean sign consists of a *representamen* for a tangible or intangible *object* that invokes an *interpretant* of the object's meaning or affect in the mind of the interpreting consumer (see Figure 2). This interpretant always arises from the percept in the subjective context or *ground* of the consumer's collateral experience external to the sign itself (Benyon-Davis, 2011). As noted by Dappert and Farquhar (2009), no unit of digital content is inherently significant; it gains significance for its consumer only "in a context relevant to some purpose or goal."

For purposes of analysis, it is useful to consider digital content in terms of five distinct semiotic dimensions: *semantics*, *syntactics*, *empirics*, *pragmatics*, and *dynamics*:[4]

1. Semantics is concerned with the relationships between content and its underlying abstract meaning or affect;
2. Syntactics, with the relationships inherent to the content's expression, that is, its symbolic encoding

[4] The tri-fold division of semiotic analysis into syntactics, semantics, and pragmatics was proposed by Morris (1946). The addition of empirics as a valid semiotic concern was suggested by Stamper (1973). Temporal considerations are usually discussed in the semiotic literature in terms of the Saussurean opposition of synchrony and diachrony (Saussure, 1916). While *chronemics* is the study of the role that time can play *as* a non-verbal sign, it generally does not consider the potentially corrosive effect of time *on* a sign (Walther and Tidwell, 1995). The alternative term used here to describe that effect, *dynamics*, is borrowed from Cheney et al. (2001) and Flouris and Meghini (2006).

structures;

3. Empirics, with the relationships between content and its physical representations;
4. Pragmatics, with the relationships between content and its consumers, that is, the executable behaviors that enable meaningful engagement; and
5. Dynamics, the relationships between various states of content as it persists and evolves across space and time.

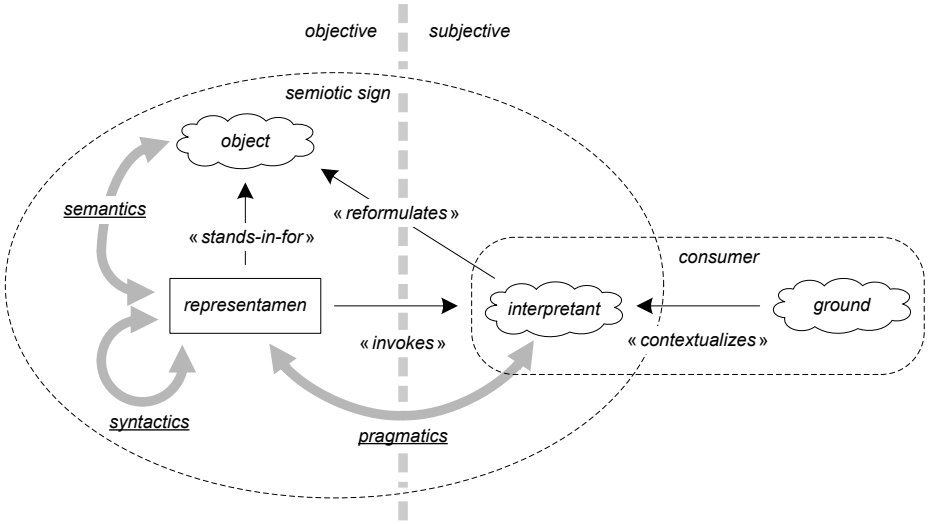

Figure 2 – Triadic semiosis

These semiotic dimensions are direct analogs for longstanding curation concerns with digital content's abstract *meaning*, symbolic inner *structure* and outer *form*, physical *carrier*, experiential *behavior*, and spatio-temporal *persistence*. The UC3 domain model name, "Sept", is derived from the approximate phonetic pronunciation of the SSEPD acronym formed by the five semiotic dimensions.[5]

2.3 Digital objects

Since the publication of Kahn and Wilensky's seminal work (1995) it has been conventional to refer to the nominal unit of digital curation analysis as a "digital object."[6] However, their definition of object – an aggregate of typed *data* and *key-metadata* – is coarse-grained and doesn't support nuanced distinctions conducive to comprehensive modeling. The OAIS reference model (ISO, 2012) generalizes and formalizes the concept of metadata as *representation information* and adds the key insight that an *information object* is an aggregate of data and the

[5] A *sept* is a division or subgroup of a Scottish family or clan, possibly derived from the Latin *sæptum*, "enclosure," which is apropos for a model defining the various affinities between related but distinct components of digital content objects.

[6] The semiotic "object" is so named in view of its position relative to the objective/subjective divide. This distinction, however, generally has not been considered in the digital library literature, where "object" is used merely as a suitably generic designation for a content carrier. Part of the impetus behind UC3's modeling effort was to provide a more precise and nuanced definition of the highly overloaded term.

representation information that defines the context in which the data are interpreted; furthermore, that representation information is itself an information object subject to its own representation information, ad infinitum. As a practical matter, the infinite regress of meta-interpretive representation information stops when it reaches the level of the presumed *knowledge base* of a *designated community*, that is, the common baseline understanding that can be assumed on the part of targeted consumers. This is an aggregate community-wide equivalent to the individual consumer's contextualizing semiotic ground.

2.3.1 Annotation

The essence of representation information is that it is an *external* description or *annotation* of its referent. Some amount of annotation – minimally, *identity* and *type* – is necessary for any effective consumption of a given unit of content. Identity serves a three-fold purpose:

1. To *demarcate* and *distinguish* a particular subset of interest from an otherwise undifferentiated digital carrier;
2. To provide a *designation* for common reference; and
3. To enable *retrieval* and *manipulation* of the identified thing.[7]

Type is necessary to establish the proper context for interpreting the syntactics of otherwise opaque digital content and thus enable the subsequent understanding of the expressed semantics and realized experience of the underlying meaning or affect. Thus, without an assertion of identity, there is no distinguishable and actionable object upon which to focus curation attention; similarly, without an assertion of type, there is no effective means of interpreting and exploiting an identified object. Content typing provides useful information at both the *individual* and *class* level. For example, the general type of *still image* implies a common set of generic attributes entailed by that type, for example, height, width, bit depth, color space, etc., while the specific JPEG 2000 format implies a particular form of syntactic structure for expressing the generic characteristics. Both the general and specific type characteristics form an important component of the preservable *essence* or *significant properties* of the represented content (Hedstrom and Lee, 2002; Dappert and Farquhar, 2009). While identity and type are fundamentally *necessary* annotative properties, by themselves they may not be fully *sufficient* for ensuring a successful engagement with identified and typed content; a fully meaningful engagement may be dependent upon higher-order semantic and pragmatic properties (APARSEN, 2014).

An annotation is a propositional statement that declares a *value* for a *property* held by its referent (Dappert and Farquhar, 2009).[8] Referents can be defined as arbitrarily-granular subsets (proper or otherwise) of a given unit of content. Collectively, annotations define the context and significance of their referents. The relationship between

[7] The three functions of identity are analogous to Dewey's characterization of a (linguistic) sign as a *fence*, separating a particular thing from all other not-things; a *label*, for common reference to the same thing by various actors; and a *vehicle*, a means to exploit the thing for some purpose (Dewey, 1910).

[8] Propositional statements are well known in the *subject-predicate-object* form used by semantic expression languages such as RDF. For example, the proposition ⟨*Jack-and-Jill*⟩ ⟨*ran-up*⟩ ⟨*the-hill*⟩ asserts that it is true that its subject ("Jack and Jill") satisfies its predicate ("ran up") relative to its object ("the hill").

an annotation and its referent content foreshadows that between the contextual ground and interpretant: annotations contribute to content's *objective* context and inform the contextualizing ground of the consumer's *subjective* interpretation (see Figure 3).

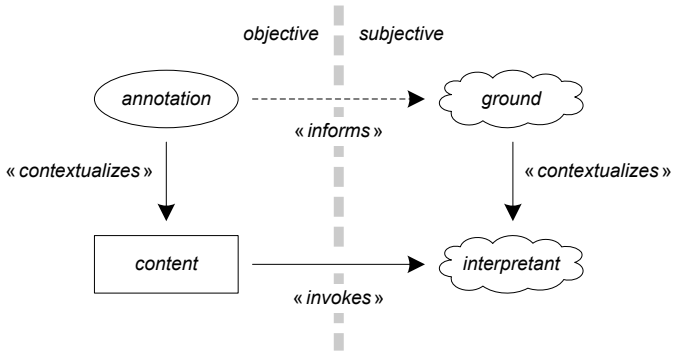

Figure 3 – Annotation and ground

Annotations can be contributed by three classes of actors: content producers, managers, and consumers; and at three points of engagement: acquisition/creation/modification, management, and retrieval/reuse. Historically, memory institutions have delegated descriptive priority to content producers and managers while discounting the role of consumers. Over archival timespans, however, the most useful content annotations may be contributed by those consumers who affirmatively seek out and exploit content, often in novel ways unintended or unforeseen by its producers. While the fullest possible set of annotations is highly desirable for facilitating successful curation strategies and outcomes, in practice this often will not be case. While this situation is not ideal, it is also not unusual: scholars have always labored in archives to (re)identify and (re)contextualize otherwise unknown or underutilized material.

2.3.2 Context and ground

Traditional semiotic analysis presupposes two primary actors in the communicative act: producers and consumers. When dealing with scholarly content in the curation domain, however, a third managerial actor generally intermediates between producer and consumer. Content is collected, managed, and presented for use by a curatorial agency as part of larger aggregations based on collection policy, thematic or disciplinary unity, or seemingly arbitrary administrative convenience or serendipity. The contextual perspective of the curatorial manager will inevitably leave traces in a consumer's interpretive experience, just as a producer's conceptual frame-of-reference informs the intention underlying content production. Thus, by virtue of its mode of production, membership in encompassing curatorial collections, and under the imprimatur of its stewardship agency, a given unit of content is situated within a wider network of explicit and implicit denotative, connotative, and metaphorical connections through which it accumulates additional nuanced meanings or affects (Danesi, 2003). These connections can take several forms:

1. Intangible associations, that is, inferences drawn from cultural and professional engagement.
    a. Reputational characteristics associated with the content producer.

b. Reputational characteristics associated with the curatorial agency stewarding collections and content.

2. Inferable relationships, that is, those that are partially or provisionally recoverable.

   a. Conceptual frames-of-reference of content producers, as reflected in producer-contributed annotations.

   b. Individual content and aggregate collection-level glosses of content managers, as reflected in manager-contributed annotations.

3. Tangible relationships, that is, those directly representable in content objects or object management systems.

   a. Typed structural and semantic relationships between separate but dependent content objects.

   b. Typed structural and semantic relationships between aggregate collections, inclusive of attendant goals, policies, interpretive context, etc., and their subsidiary member objects.

Thus, the consumer's subjective ground is a complex admixture influenced by originating frame-of-reference, intention, annotation, and association; curatorial context, selection, arrangement, annotation, and association; and personal experience, expertise, and expectation (see Figure 4).

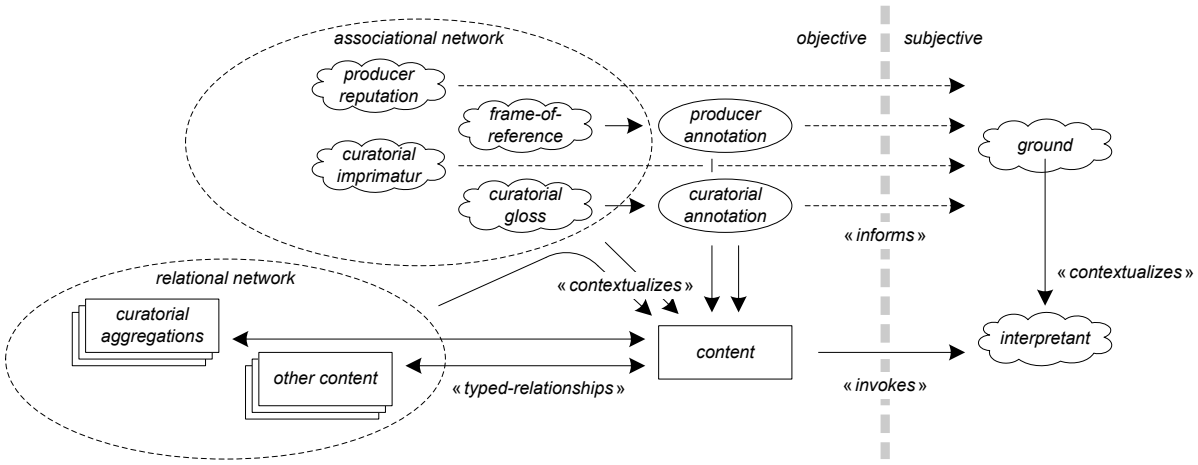

Figure 4 – Context and ground

The set of connections in which a given object participates form *associational* and *relational networks* that are fundamentally fluid and dynamic, as every act of dereferencing has the potential to result in modification or addition, particularly to the intangible associational network (Steels, 2008). The generalized recursive nature of these networks – the object of any given connection is itself susceptible of further connection – is reminiscent of the chains of references in OAIS (2012) representation networks.

2.3.3 Behavior

A human consumer never directly interacts with a digital object in its native digital form. This realization underlies the NAA performance model (Heslop et al., 2002), which stresses the importance of an object's *performance*, that

is, its transformation into humanly-sensible form. A performance is equivalent to what Doerr and Tzitzikas's information carrying ontology (ICO) model calls a *projection*, a process for converting an information carrier into a sensory impression (Doerr and Tzitzikas, 2012). The KB preservation layer model (PLM) introduced the term "view path" to refer to the complete set of technical dependencies by which a digital object is *rendered* into analog form (van Diessen, 2002).[9] These dependencies exist in the context of a larger computational *environment* including software applications, operating systems, and hardware platforms, with each level being susceptible to more granular decomposition. The importance of documenting view paths cannot be overstated since *all* engagement with digital content entails intermediation through executable behaviors to convert the content to an apprehendable form. It is preferable to speak of "behaviors" rather than "view paths" as the latter carries an implication of being applicable only to passive consumption (i.e., *viewing*) while the former better encompasses arbitrarily-complex engagement modalities, for example, interactive, dynamic, immersive, etc.

2.3.4 Materiality

All of the modeling efforts discussed so far start from the premise that a digital object is a *physical* thing, that is, an organized pattern of matter or energy in the sense described by Bates (2006). The lowest stratum of OAIS data modeling, for example, is defined as a set of bits. The FRBR model provides a valuable insight that content exists – or can be construed – along a continuum between intangibility and materiality: an abstract *work* realized by a symbolic *expression* physically embodied in a *manifestation* exemplified by an *item* (IFLA, 1998). A similar hierarchy underlies the basic representation model (BRM), in which abstract *propositional content* is expressed by nested *symbol structures* that are inscribed in *patterned matter and energy* (Wickett et al., 2012). The "propositional content" terminology, however, may be too limiting for a general-purpose curation model since it implies that the BRM is applicable only to content reducible to factual truth claims.[10] In order to encompass non-factual content, for example, creative or aesthetic works intended to impart a direct visceral emotional response, the more generic label "noumenon" is preferable as it applies to the widest possible range of content types.[11] In the semiotic triad, a noumenon plays the role of the semiotic object. The noumenon, its encoded inscription on a physical carrier, and the behaviorally-mediated engagement with it correspond to Buckland's *information-as-knowledge/information-as-thing/information-as-process* trichotomy (Buckland, 1991).

2.3.5 Dynamics

The goal of curation is to provide a stable interpretive *experience* of a given noumenal unit of content across space and time. Ubiquitous connectivity has largely eliminated difficulties posed by spatial concerns, but the inherently corrosive effects of time on digital content remain problematic. As Moore (2008) and Mois et al. (2009) have

---

[9] Although the term "rendered" is most commonly applied to images, it is used here in a generic sense of *any* digital-to-analog conversion regardless of content genre or format.

[10] As the BRM was developed to model scientific data, that is, "all and only those things that are either possibly true or possibly false" (Wickett et al., 2012), this terminology is appropriate in its original context.

[11] "Noumenon," from the Greek νοούμενον, "that which is apprehended by thought," in distinction to *phenomenon*, "that which is apprehended by sight," is used by Kant (1781) to refer to a posited thing-in-itself (*das Ding an sich*) independent of any material representation.

pointed out, digital preservation (and by implication, curation) is "communication with the future." The seminal model of information-theoretic communication was developed by Shannon (1948), in which a *sender* encodes a *message* and transmits it as a physical *signal* across a *channel*, subject to degrading *noise*, at which point it is decoded and experienced by a *receiver*. In subsequent refinements, Schramm (1954) emphasized the necessity of the (human) sender and receiver sharing a common *frame of reference* for the communicative act to be successful, while Berlo (1960) introduced consideration of the *effect* the communicated message has on its receiver (see Figure 5). The frame-of-reference is equivalent to the assumption of *a priori* knowledge of an information object's symbol set and arrangement rules that is a necessary precondition for its effective exploitation in the ICO model (Doerr and Tzitzikas, 2012), both of which inform the subjective semiotic ground.

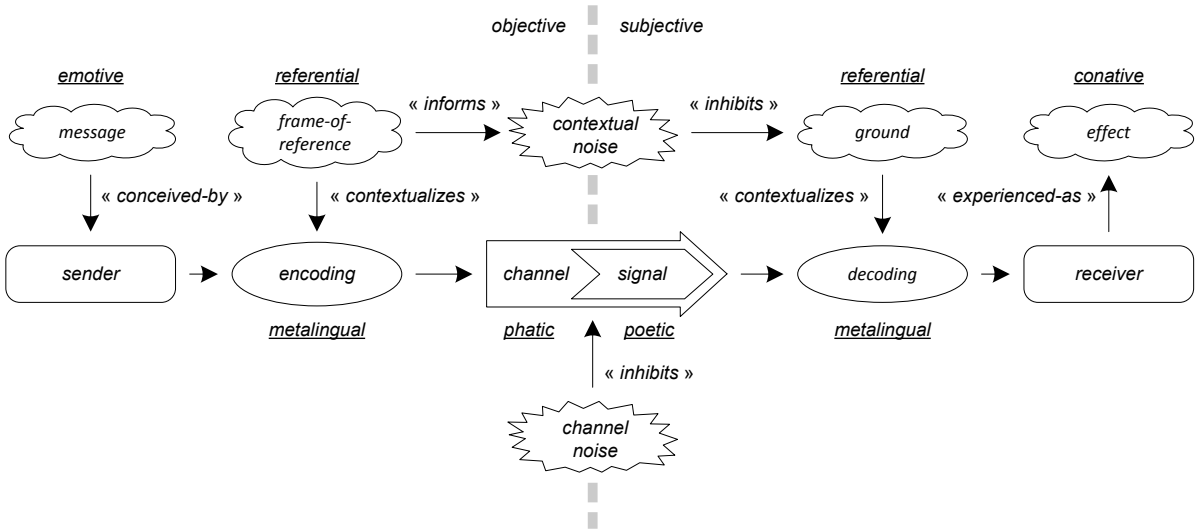

Figure 5 – Sender/receiver communication model

Factors that impede communication are *noise*. It is useful to distinguish between *channel* and *contextual* noise: the former degrades the integrity of the signal carrying the encoded message – for example, media bit corruption introduced through proximity to ionizing radiation or magnetic fields, thermal stress, mechanical trauma, etc. – while the latter distorts the interpretive context of the message – for example, a conceptual misalignment between the objective frame-of-reference and subjective contextual ground – and thus, the message's interpretation and ultimate effect on its receiver. The primary strategy for ameliorating the effects of channel noise is the addition of *redundancy* to the encoded signal, for example, parity, checksums, erasure codes, mirroring, etc. One strategy for minimizing contextual noise is to provision effective means for the sender's frame-of-reference to inform fully the contextual ground of the receiver; in other words, to ensure that the receiver can properly recover the *intention* of the sender. Descriptive annotations are included as a fundamental component of a digital object in order to facilitate this very process.[12] However, since this strategy implies communication of the

[12] The theory of the *intentional fallacy* holds that it is never possible to be fully cognizant of the intention of a message's sender, since there is no reliable means extrinsic to the message itself for the receiver to determine that intention (Wimsatt and Beardsley, 1946). In other words, from the perspective of a content consumer only interpretation can be known definitely, while intention can be intuited – at best – only provisionally. However, while clearly appropriate to inherently

annotations across a channel either in conjunction with, or independent of, their referent content, the amelioration of contextual noise is itself subject to potential channel noise. The message, frame-of-reference/context, encoding/decoding, channel, signal, and effect components of the augmented sender-receiver model correspond respectively to the *emotive*, *referential*, *metalingual*, *phatic*, *poetic*, and *conative* functions of (linguistic) communication identified by Jakobson (1960).

A curation channel may have both spatial and temporal extension and thus an object's state is inherently situated within a digital space-time frame. Cheney et al. (2001) view the dynamics of a digital object in game-theoretic terms as a sequence of state configurations resulting from alternating game moves or *actions* by the object's encompassing environment and curating agent, respectively introducing inhibiting channel and contextual noise and applying both proactive and reactive ameliorating strategies. The interactions between a responsible curating agent and malicious actors seeking to subvert curation goals for their own gain can be characterized as a *competitive* game (Chatain, 2013), in which one side's advantage is the other's loss. On the other hand, interactions with encompassing physical and computational environments are neither strictly competitive nor *cooperative*, since the environmental actions are impersonal and undirected towards any particular teleological purpose.

2.3.6 Object model

Synthesizing these various formulations, a digital object can be defined as consisting of (see Figure 6):

1. An essential noumenal meaning or affect, ...
2. Expressed via nested symbolic encodings, ...
3. Inscribed on physical carriers, ...
4. Described by annotations (themselves objects), ...
5. Realized through behaviors (themselves objects); and ...
6. Subject to actions.

That is, an ineffable noumenon is given *abstract* form through various symbolic encodings, *physical* form on a digital carrier, experiential *reality* through mediating behaviors, and *persistence* through intervening actions. A noumenon's complete encoding may result in a highly composite entity composed of multiple structurally-related files and hierarchically-nested bitstreams, each with some degree of independent identity and function, but nevertheless contributing to a larger coherent whole. In this sense a noumenon's encoded expression is equivalent to a PREMIS *representation* (PREMIS, 2012). Note that annotations and behaviors are themselves representation information-like objects, which can be incorporated by encapsulation (perhaps most appropriate for annotations) or reference (most appropriate for behaviors). The latter case relies on an external curation environment to ensure that the referents are available as necessary. Behaviors correspond to service definition

---

subjective creative or non-propositional content (Morrissey, 2014), the intentional fallacy may not apply to content comprised of (purportedly) factual truth-claims.

and deployment objects in the Fedora object model (2012).

While the conceptual modeling of a digital object progresses most naturally in a top-down (or abstract-to-tangible) direction, that is, noumenon > encoding > carrier, as a practical matter it is useful to consider the bottom-up progression from most general to most specialized organization, carrier > encoding > noumenon, since curation strategies focused on generic objects usually apply equally well to specializations, but not necessarily vice versa. Thus it is useful to distinguish six hierarchical stages of accumulating object utility: *entities*, *artifacts*, *articles*, *commodities*, *assets*, and *heirlooms* (see Figure 7 and Table 1).[13]

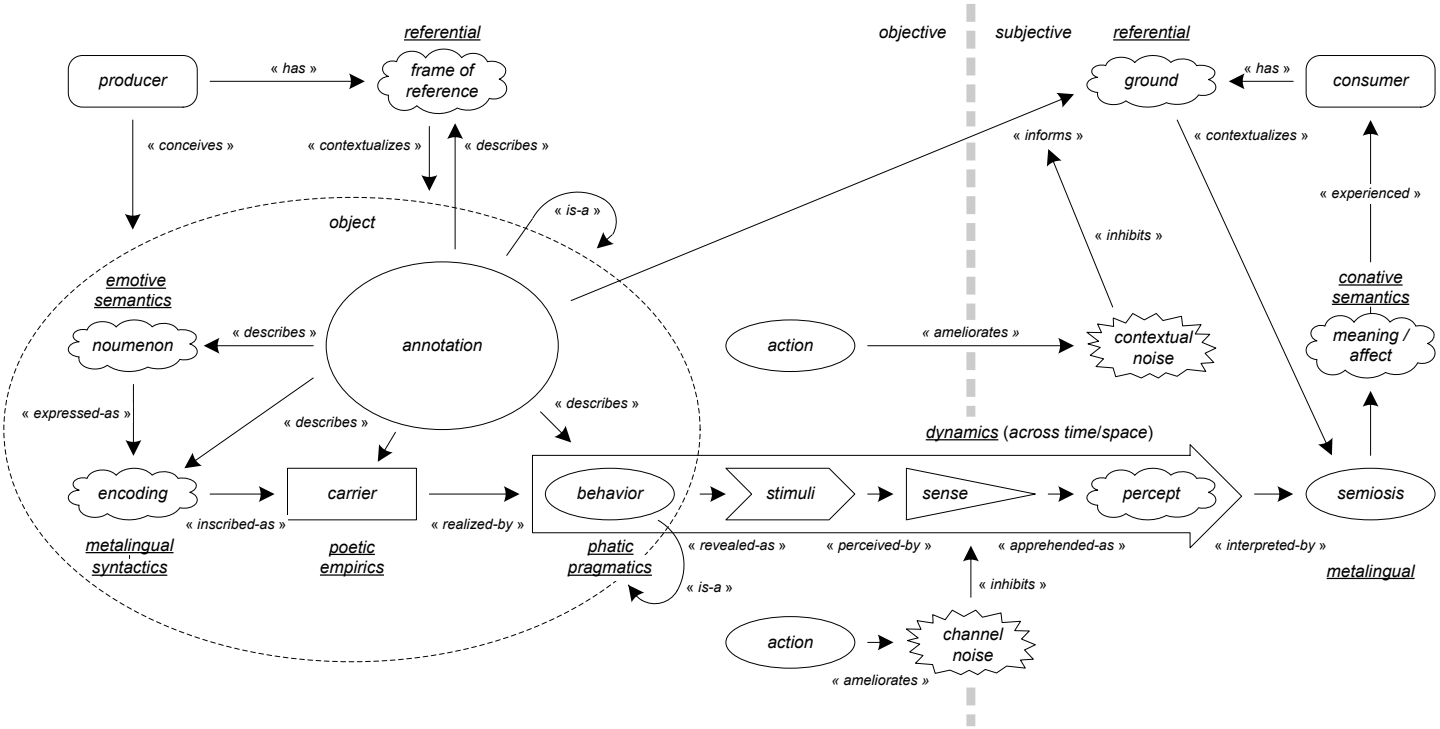

Figure 6 – Curation object model

1. An entity is a digital *thing*. It comes into being through a formative act of *inscription* that results in tangible bits on an otherwise undifferentiated digital carrier. Beyond its own existence nothing further can be known or inferred about a digital entity. Thus, its essential quality is *existential* and its utility is *nascent*. For example, consider the *bits* …00000000000000000000000000011000110101001010000… inscribed on a digital carrier, which by themselves convey no useful information.
2. An artifact is an *intentional* entity, that is, one can infer that it was deliberately created, even though the purpose underlying that creation remains undisclosed. An artifact comes into being through an informative act of *identification* that demarcates a *particular* sequence of bits fixed in digital space-time.

[13] Although there are only five semiotic dimensions it is useful to distinguish between *inner* and *outer* symbolic encodings, corresponding respectively to commodities and articles in the six-level object typology, but both associated with the syntactic dimension.

This constitutes an *outer* symbolic encoding concerned with the artifact's external morphology, that is, its interface with its encompassing computational environment, for example, encoding details attendant to tracks, sectors, blocks, volumes, file systems, etc. Many artifacts will possess a number of hierarchically-nested encodings. The essential properties of an artifact are its name, location, size, bitstream, and any function over that bitstream, for example, a message digest, that is not dependent on higher-order syntactic, semantic, or pragmatic information. In and of itself, however, an artifact is syntactically opaque: it affords no opportunity to interpret or infer how its constituent bits express any underlying noumenal meaning. Thus, its utility is *incipient*. For example, consider the *file* "E20140714112409.jp2" with specific size, creation and last modification date/times, MD5 digest, etc., absent any knowledge of its contents' expression.

3. An article is a *purposeful* artifact, even if the meaning underlying that purpose remains undisclosed. An article comes into being through an informative act of *characterization* that documents *inner* structural symbolic encodings concerned with the internal structural expression of the article.[14] The essential properties of an article are its type or format and any further attributes entailed by that format. Many articles will possess a number of hierarchically-nested encodings. While these provide details of the article's expression, its underlying noumenal meaning is still semantically opaque. Thus, an article's utility is *potential*. For example, consider the JPEG 2000-formatted *image* with three 8-bit components representing sRGB color samples, with 1024-x1024 tiles, 64x64 code blocks, six decomposition layers, 25 quality layers, and 9-7 irreversible wavelet compression, absent any knowledge of what the image *represents*.

4. A commodity is a *meaningful* article. It comes into being through an informative act of *description* that documents its underlying noumenal meaning or affect in terms common to an appropriate domain of scholarly discourse. In and of itself, however, a commodity doesn't afford any practical means to experience or exploit that meaning or affect. Thus, a commodity's utility is *theoretical*. For example, consider the photographic image of Lake Merritt, a national historic landmark and the United States' first designated wildlife refuge located at 37.8039° N, 122.2591° W, close to UC3's offices in Oakland, California, absent any revealing behaviors.

5. An asset is a *useful* commodity. It comes into being through a performative act of *realization* that exposes the article's meaning or affect as stimuli apprehendable to human sensory modalities.[15] An asset's utility is *practical*: it can be directly experienced and exploited by a consumer towards some useful purpose. For example, consider a consumer's *experience* engaging with the Lake Merritt image in a colorimetric image processing environment supporting dynamic zooming, panning, cropping, annotation, etc., absent any consideration of spatial or temporal extension.

6. An heirloom is a *reliable* asset. It comes into being through a reformative act of *intervention* that ensures

[14] The inner/outer distinction is borrowed from Hofstadter (1979), although with a somewhat different emphasis.

[15] The act of engagement underlying an asset presupposes syntactics- or semantics-aware behaviors. Merely being able to provide file-level access to or manipulation of an artifact is not sufficient to raise its status to an asset.

the continuing viability and usability of the asset across space and time. To the extent to which those interventions are successful, an heirloom's utility is *enduring*. For example, consider a consumer's *future* engagement experience with the Lake Merritt image.

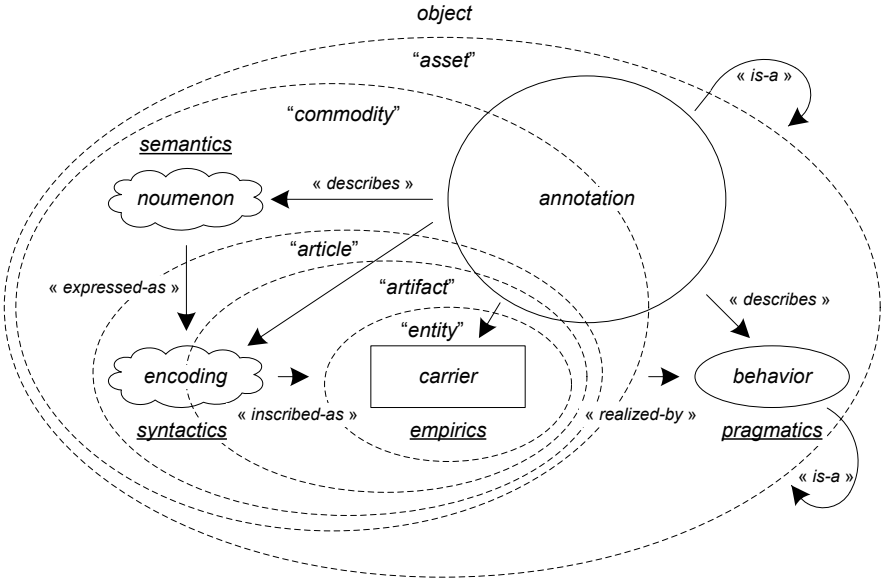

Figure 7 – Curation object typology

Artifacts, commodities, assets, and heirlooms correspond respectively to the *identity*, *understandability*, *renderability*, and *persistence* concerns of the simple property-oriented threat (SPOT) model (Vermaaten et al., 2012); the SPOT *availability* and *authenticity* properties are cross-cutting concerns applicable across the typological hierarchy. The particular arrangement of inner and outer symbolic encodings expressing a given article may conform to a common pattern, or *content model* (Fedora Commons, 2011). Classifying articles by their content model may aid curation activities by permitting behaviors, analyses, policies, strategies, and actions to be defined at a class rather than individual level.

Differentia	*Entity*	*Artifact*	*Article*	*Commodity*	*Asset*	*Heirloom*
Dimension	*empirics*	*syntactics*	*syntactics*	*semantics*	*pragmatics*	*dynamics*
Mode	*formative*	*informative*	*informative*	*informative*	*performative*	*reformative*
Act	*inscription*	*identification*	*characterization*	*description*	*realization*	*intervention*
Concern	*media*	*form*	*structure*	*meaning/affect*	*experience*	*persistence*
Abstraction	*carrier*	*(outer) encoding*	*(inner) encoding*	*noumenon*	*behavior*	*action*
Quality	*existential*	*intentional*	*purposeful*	*meaningful*	*useful*	*reliable*
Utility	*nascent*	*incipient*	*potential*	*theoretical*	*practical*	*enduring*
Annotation	*provenancial / administrative / permissive*	*morphological / relational / associational*	*structural*	*intellectual*	*instrumental*	*provenancial*

Table 1 – Curation object typology

Table 1 indicates an association between the six typological classes and a set of annotation categories: *provenancial*, *administrative*, *permissive*, *morphological*, *structural*, *intellectual*, and *instrumental*:

1. Provenancial annotations describe the actors, conditions, and events that led to the creation, acquisition, or revision of the content;
2. Administrative annotations describe the actors, conditions, and events related to the ongoing curation management of the content;
3. Relational annotations describe structural connections with other objects and aggregated collections.
4. Associational annotations describe producer frames-of-reference and curatorial policies and interpretive glosses.
5. Permissive annotations describe IPR and terms of service rights and obligations attendant to content management and engagement;
6. Morphological annotations describe content's externally-facing expression in terms of outer symbolic encodings;
7. Structural annotations describe content's internal expression in terms of inner symbolic encodings;
8. Intellectual annotations describe content in terms meaningful to an applicable domain of scholarly discourse; and
9. Instrumental annotations describe behaviors applicable to the content.

This is not to imply that these categories are appropriate *only* relative to their associated classes; rather, the associations merely indicate the earliest stage in the typological progression at which those particular annotation types are relevant. For example, an entity has provenancial properties independent of and prior to any artifactual concerns (for example, carrier *X* was received from agent *Y* at time *Z*, etc.), an artifact has morphological properties independent of and prior to any article-level concerns (file *X* of size *Y* and modification date *Z*, etc.), and so on.

Neither the model components nor its typological classes should be construed as clearly-demarcated categories. Rather, they represent useful abstractions whose properties, coalescing around core conceptual centers of gravity, may be held to a greater or lesser extent by any particular component or typological instantiation. Nevertheless, the components and typology can be used to make concise yet precise statements regarding curation policies, strategies, systems, and results. For example, it is common to divide preservation obligations into tripartite media, bit-level, and functional preservation levels. These correspond respectively to activities focused on ensuring the integrity of entities, articles, and assets. Creating forensic disk images is a suitable strategy for preserving entities (that is, media objects), independent of any artifactual morphology; fixity audit, for artifacts (file objects), independent of any type characterization; migration, for commodities (syntactically- and semantically-characterized objects), independent of any behavioral considerations; and emulation, for assets (experiential objects).

While a curating agent could choose to enforce a lower service obligation than what may be otherwise

supportable by an object's typological characteristics, it is not possible to meet a higher obligation. For example, an article (e.g., a typed file) could be managed purely as an artifact (an opaque file) through the expedient of disregarding any non-morphological characterization, but no matter how successful the preservation of a true artifact, it will never afford any higher-order structural information about its contents; if such information were known or could be inferred, the object would be an article rather than an artifact. Thus, finely-grained typological modeling permits more precise statements of curation intention, expectation, and result. For example, saying that an object will be "functionally" preserved is open to potential ambiguity; on the other hand, saying that it will be preserved as an article makes clear that it will continue to be a purposeful object through persistent association with pertinent inner structural encoding information. Similarly, a preserved commodity will remain meaningful through association with appropriate semantic characterization, and a preserved asset will remain useful through association with realizing behaviors.

Given a semiotic view of content engagement, it may never be possible to preserve a digital object "perfectly." While it is potentially possible to fix and maintain indefinitely the state for all components on the objective side of the communication channel (i.e., noumenon, encoding, carrier, annotation, behavior, stimuli), on the subjective side the consumer's future contextual ground, conditioned by the totality of the consumer's intervening lived experience, is not susceptible to any equivalent constraint. This may not be significant for propositional content consisting of purportedly-objective truth claims, but could be important for creative content.

While it is often convenient to refer of a digital object's state as unitary, each individual object component – encoding, carrier, annotation, behavior, etc. – can evolve independently along its own unique dynamic trajectory.

2.4 Modeling content engagement

An act of consumer engagement occurs when a digital object is:

1. Acted upon by *behaviors*, …
2. Rendered as physical *stimuli*, …
3. Apprehended as a *percept*, …
4. Reformulated as an *interpretant* in the subjective context of the consumer's *ground*, and is …
5. Experienced as subjective *meaning* or *affect*.

Content engagement is modeled in terms of a set of *actors* who engage with digital content in various ways across a scholarly and curatorial *life-cycle* (Ball, 2012; CEOS Working Group on Information Systems and Services, 2012), and the life-cycle *activities* in which those actors participate. At the highest level of abstraction, there are three categories of actorial roles, although each can be usefully discriminated into more granular parts:

1. Content *producers*, actors who create or acquire digital content and exercise *ultimate* intellectual, administrative, financial, and permissive *control* and *responsibility* for its ongoing stewardship;
2. Content *managers*, actors who oversee managed content and exercise *delegated* administrative,

technical, and instrumental control and responsibility for its ongoing stewardship; and

3. Content *consumers*, actors who *exploit* managed content for some purpose.

These roles correspond to the producer, archive/management, and consumer entities in the OAIS reference model (ISO 14721, 2012). "Actor" should be construed in the most general sense as either a corporate or human agent; even if they make extensive use of technological facilities, ultimate curation agency always resides in institutional or individual actors. There are three high-level life-cycle stages corresponding to the main activities of these actors; again, all are amenable to more nuanced internal distinctions:

1. Content *production*;
2. Content *management*; and
3. Content *consumption*.

As Upward points out (1997), however, it may be more useful to talk about an activity *continuum* rather than life-cycle, as the later implies a linear progression through well-defined and clearly demarcated stages. In distinction, a continuum approach emphasizes the essential non-linear contiguity and overlapping interdependence of curation activities. Thus, it is appropriate to group various curation engagement activities by thematic loci within a permeable continuum, characterized by dimensions of first-order *origination*, concerned with establishing the resource objects of curation focus; second-order *organization*, concerned with codifying and imposing appropriate structure on curation resources; and third-order *pluralization*, concerned with expanding the reach and consequence of curation resources, rather than discrete life-cycle stages (see Table 2). A comprehensive curation program will formulate – and over time, maintain, revise, and extend – controlling policies, executable strategies, and implementing guidance, systems, services, and operational procedures for all major continuum resources and activities.

Locus	*Originate*	*Organize*	*Pluralize*
Production	*capture, simulate, create, derive*	*identify, classify, clean, annotate, package*	*license, submit, publish, cite, aggregate*
Management	*appraise, select, harvest, collect*	*characterize, normalize, store, index, plan, watch, intervene, administer*	*replicate, audit, notify, syndicate, report*
Consumption	*search, discover, retrieve, subselect*	*analyze, synthesize, correlate, interpret, transform, annotate*	*summarize, assert, validate, refute*

Table 2 – Curation engagement continuum

2.5 Modeling curation policies and strategies

A formal statement of curation policy is necessary to set expectations properly and form the basis for acceptable terms of service and assessment of the efficacy of curation outcomes. Strategies represent specific organizational intensions for fulfilling or enforcing promulgated policies (Becker at al., 2009; Sierman, 2014). Strategies are implemented by concrete plans and activities. Curation policies, strategies, plans, and activities are modeled in terms of a set of high-level categories of consideration. These consist of one preparatory imperative:

0. *Predilect*. Decide what you intend to do.

and five implementation imperatives:

1. *Collect*. Obtain what you intend.
2. *Protect*. Preserve what you obtain.
3. *Introspect*. Know what you protect.
4. *Project*. Offer what you know.
5. *Connect*. Deliver what you offer.

Although with most obvious applicability to content, these imperatives are equally relevant to other aspects of the curation complex, for example, technical infrastructure, operational procedures, staffing, etc.

Each implementation consideration can be applied to every typological class (see Table 3). While the resulting matrix is suggestive of the NDSA levels of preservation (Library of Congress, 2013; Philips et al., 2013), the typological progression (entity, artifact, etc.) plays a different role than the NDSA levels as it is defines accumulating levels of utility rather than assurance. There is a general inheritance of relevant considerations across the typological and imperative progressions.

Imperative	*Entity*	*Artifact*	*Article*	*Commodity*	*Asset*	*Heirloom*
Predilect	*service level agreement*	*disaster recovery / business continuity*	*action plans*	*collection development policy*	*outreach*	*sustainability and succession planning*
Collect	*annotation*	*packaging, submission*	*normalization / canonicalization*	*workflow / tool integration*	*code / workflow repository*	*chain of custody*
Protect	*environmental control, write blocking, media refresh, redundancy, decorrelation*	*administrative control, malware detection / sanitization, heterogeneity, fixity audit*	*technical control, migration*	*bibliographic control*	*access control, emulation*	*change control, technology watch*
Introspect	*forensic characterization and annotation*	*morphological characterization and annotation, identifier minting*	*structural characterization and annotation, ontologies, technical registry*	*intellectual characterization and annotation, entity extraction, sentiment analysis, master registry, identifier binding*	*behavioral characterization and annotation, software registry, analytics*	*provenance, annotation*
Project	*media inventory*	*file inventory, identifier resolution*	*object inventory*	*catalog*	*transcoding, syndication, aggregation, discovery, outreach, training*	*versioned change history*
Connect	*legacy / emulated computational environments*	*file delivery*	*local format-aware manipulation*	*local domain-aware manipulation*	*search / browse, hosted manipulation, annotation*	*consortial collaboration*

Table 3 – Curation policy and strategy

A media inventory, file inventory, object inventory, and catalog are distinguished by respectively exposing for public view and retrieval enumerations of physical media, (opaque) files, aggregations of structurally-related and/or constrained files and bitstreams, and intellectually-coherent works.

The foundational imperative for all curation activities is *collection*, that is, bringing content into an appropriate technical infrastructure under the aegis of a responsible managerial actor. While it is *possible* that collected content will not be fully susceptible to curation strategies leading to successful outcomes, it is almost *certain* that uncollected content will result in curation failure (Rosenthal, 2014). The baseline level of preservation assurance that can be realistically-asserted by a responsible curating agent is either as an entity or artifact, depending upon whether the content was collected as (undifferentiated) media or (opaque) files. Increasingly higher-order outcomes may be possible if the collected content meets the incrementally more stringent criteria for articles, commodities, assets, or heirlooms, that is, the content encapsulates or is accompanied by appropriate characterizing syntactic, semantic, or pragmatic annotations, and is subject to appropriate dynamic intervention to combat degradation across space and time.

The notion of a range of possible outcomes is fundamental to UC3's stated preservation obligation:

> CDL aims to provide the highest level of preservation for User's content, as defined by commonly-accepted community standards and best practices, that is consistent with the form, structure, and packaging of the managed digital content, the degree to which that content is accompanied by authoritative and comprehensive metadata, the availability of appropriate tools, and organizational priorities. Note that this implies a continuum of preservation outcomes dependent on the nature of the content submitted by User, although at a minimum, CDL is committed to providing bit-level preservation of all content (CDL, 2014).

This public statement of policy can be recast more formally in the technical language of the UC3 Sept model:

1. UC3 will accept custodial stewardship for arbitrary digital content regardless of its position in the typological spectrum from artifact to asset.
2. UC3 does not (currently) accept custodianship of entities (i.e., physical media objects).
3. UC3's custodial obligation is to preserve content to the extent consistent with its typological basis, that is, ensuring the:
    a. Morphological integrity of artifacts (i.e., opaque objects);
    b. Structural integrity of articles (syntactically-characterized objects);
    c. Semantic integrity of commodities (semantically-characterized objects); and the
    d. Behavioral integrity of assets (pragmatically-characterized objects).
4. UC3 will provide effective technical mechanisms to acquire and manage the annotations necessary to support a given object's typological integrity.
5. UC3 will provide effective technical mechanisms to acquire, manage, and execute the behaviors necessary

to realize a given asset's pragmatic integrity.

3 Curation architecture and infrastructure

A model represents a template or design that must be implemented in order to achieve its desired goals. The information technology community has long used the metaphor of *architecture* to refer to the large-scale organization of software, hardware, and networked systems, drawing inspiration from principles long associated with the built environment. The architectural ideals promoted by the Roman theorist Vitruvius (1634) – commodity, firmness, and delight – can be usefully transformed into cognate technological characteristics of *utility*, *resilience*, and *elegance*; that is, architectural components should fully meet some well-defined need; be accepting of the widest possible range of input conditions and tolerant of failure; and display ingenuity and economy of design, implementation, and operation. Architectural design, implementation, and operation should be guided by theory, informed by intuition, and tempered by experience. Individual architectural components must be tolerant of failure, responsive to scale, and adaptive to rapid ongoing, and often disruptive, evolution of customer expectations and requirements, community standards and practices, technology, and institutional mission.

4 Summary

The digital curation field has reached a stage of maturity where it can usefully draw upon a rich body of theoretical research and practical experience. Many specific segments of the curation domain have been subject to modeling activities, but the scope, coverage, and granularity of this work has varied widely. In an effort to produce a comprehensive view of the domain for purposes of analysis, planning, and evaluation of its activities, the UC Curation Center has synthesized and reformulated the many valuable contributions of prior efforts such as FRBR, OAIS, NAA, PLM, PREMIS, BRM, ICO, SPOT, and NDSA into a new inclusive model. (See Appendix A for a conceptual crosswalk between UC3's Sept and prior models.)

One important insight of the UC3 modeling effort is that engagement with digital content is an inherently semiotic process. Thus, UC3's Sept model approaches all aspects of curation activities through the lens of the five semiotic dimensions of semantics, syntactics, empirics, pragmatics, and dynamics, which correspond to longstanding curation concerns with abstract noumenal meaning or affect, inner and outer encoding structures, physical carriers, realizing behaviors, and persistence and evolution through time. This leads to a hierarchical typology of accumulating content utility consisting of entities, which are arbitrary digital things; artifacts, which are intentional entities; articles, purposeful artifacts; commodities, meaningful articles; assets, useful commodities; and heirlooms, reliable assets. Engagement with curated content is modeled by three classes of actor roles and loci of concerns – producers/production, managers/management, consumers/consumption – all within a continuum of originating, organizing, and pluralizing dimensions that address concerns of establishing, imposing order and structure upon, and extending the reach and consequence of curation resources encompassing content, activities, staff, and technology. Curation intentions, policies, strategies are modeled by six imperatives: predilect, collect, protect, introspect, project, and connect that apply respectively to considerations of analysis, decision making and planning; acquisition and collection; preservation across space and time; semantic, syntactic, and pragmatic

characterization; publication and notification; and consumer engagement.

All of the Sept model components were developed incrementally from first principles in an effort to ensure comprehensive applicability and internal consistency. While the model introduces unfamiliar terminology, UC3 believes that this vocabulary supports important nuanced distinctions in the delineation of content, content engagement, and curation policies and strategies. The model's granular definition permits the concise statement of common curation intentions, activities, and outcomes. It forms the basis for UC3's decision-making processes regarding curation infrastructure, services, and initiatives, and may be of interest to the wider curation community, with which it shares many common concerns and practices.

A Digital object model crosswalk

Table A1 provides a conceptual crosswalk between the UC3 Sept model and prior modeling efforts: sender/receiver model (Shannon, 1948; Schram, 1954; Berlo, 1960), Buckland (1991), Kahn and Wilensky (1995), FRBR (IFLA, 1998), OAIS (ISO, 2012), NAA (Heslop et al., 2002), BRM (Wickett et al., 2012), ICO (Doerr and Tzitzikas, 2012), and PREMIS (2012). Given the varied domains and methodologies underlying these models, the alignment of model components is in some cases more approximate than others.

Sept	*Sender / receiver*	*Buckland*	*Kahn-Wilensky*	*FRBR*	*OAIS*	*NAA*	*BRM*	*ICO*	*PREMIS*
noumenon	*message*	*information-as-knowledge*		*work*			*propositional content*		*intellectual entity*
inner encoding	*encoding*	*information-as-thing*	*data*	*expression*	*data object*	*essence*	*symbol structure*	*symbol structure*	*bitstream*
outer encoding				*manifestation*					*filestream*
carrier	*channel*			*item*	*digital object*	*source*	*patterned matter/energy*	*information carrier*	*file / representation*
annotation	*frame-of-reference*		*key-metadata*		*representation information*		*auxiliary information*		
behavior	*sender*	*information-as-process*				*process*		*projection*	
stimuli	*signal*					*performance*		*sensory impression*	
percept	*receiver*								
ground	*context*				*knowledge base*				
interpretant	*effect*								
entity			*digital object*		*information object*	*digital record*	*digital object*	*digital object*	*object*
artifact									
article									
asset									
commodity									
heirloom									

Table A1 – Digital object model crosswalk

Aligning the model components against one another illustrates the overall fineness of granularity with which various curation programs view digital content. It also discloses that most models do not incorporate the full set of modeled granules. The UC3 Sept model was created in part with the goal of having analogs for *all* levels of content granules comprehensively defined in a single coherent model.

Acknowledgements

The underlying Sept model and its description in this whitepaper have been materially improved through discussions with Christoph Becker, Michael Buckland, John Kunze, Jerry McDonough, and Sheila Morrissey.

References

[1] APARSEN [Alliance for Permanent Access to the Records of Science Network] (2014), *D11.3 Report on a common vision of digital preservation: Progress to year 3*, APARSEN-REP-D11_3-01-1_1, urn:nbn:de:101-2014051634 <http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2014/06/APARSEN-REP-D11_3-01-1_1_inclURN.pdf>.

[2] Baker, Alan (2010), "Simplicity," *Stanford Encyclopedia of Philosophy* <http://plato.stanford.edu/entries/simplicity/>.

[3] Ball, Alex (2012), *Review of Data Management Lifecycle Models*, version 1.0 (University of Bath) <http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf>.

[4] Bates, Marcia (2006), "Fundamental forms of information," *Journal of the American Society for Information Science and Technology* 57.8 (April 25): 1033-1045 <http://dx.doi.org/10.1002/asi.20369>.

[5] Becker, Christoph, Hannes Kulovits, Mark Guttenbrunner, Stephen Strodl, Andreas Rauber, and Hans Hoffman (2009), "Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans," *International Journal on Digital Libraries* 10.4 (December): 133-157 <doi: 10.1007/s00799-009-0057-1>.

[6] Berlo, David (1960), *The Process of Communication* (New York: Holt, Rinehart, and Winston).

[7] Beynon-Davis, Paul (2011), *Significance: Exploring the Nature of Information, Systems and Technology* (Palgrave Macmillian).

[8] Buckland, Michael (1991), "Information as thing," *Journal of the American Society for Information Science* 42:5 (June): 351-360 <http://people.ischool.berkeley.edu/~buckland/thing.html>.

[9] CDL [California Digital Library] (2014), *Terms of Service*, draft.

[10] Chatain, Olivier (2013), "Cooperative and non-cooperative game theory," *Palgrave Encyclopedia of Strategic Management*, ed. Mie Augier and David Teece <http://works.bepress.com/cgi/viewcontent.cgi?article=1017&context=olivier_chatain>.

[11] Cheney, James, Carl Lagoze, and Peter Botticelli (2001), *Towards a Theory of Information Preservation*, Cornell University computer science technical report TR 2001-1841 <http://hdl.handle.net/1813/5828>; also available as "Towards a theory of information preservation," *5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL '01)*, Darmstadt, September 4-9; also published as *Lecture Notes in Computer Science* 2163, ed. Panos Constantopoulos and Ingeborg Sølberg (Berlin: Springer-Verlag): 340-351 <http://link.springer.com/chapter/10.1007/3-540-44796-2_29>.

[12] CEOS [Committee on Earth Observation Satellites] Working Group on Information Systems and Services (2012), *Data Life Cycle Models and Concepts, CEOS Version 1.2*, version 13.0 <http://www.ceos.org/images/DSIG/Documents/Data_Lifecycle_Models_and_Concepts_v13-1.docx>.

[13] Consultative Committee for Space Data Systems (2012), *Reference Model for an Open Archival Information System (OAIS)*, recommended practice CCSDS 650.0-M-2 <http://public.ccsds.org/publications/archive/650x0m2.pdf>.

[14] Dallas, Costis (2007), "An agency-oriented approach to digital curation theory and practice," *ICHIM '07, International*

Cultural Heritage Informatics Meeting,Toronto, October 24-26, ed. J. Trant and B. Bearman (Toronto: Archives and Museum Informations): 49-72 <http://www.archimuse.com/ichim07/papers/dallas/dallas.html>.

[15] Danesi, Marcel (2003, "Metaphorical 'networks' and verbal communication: A semiotic perspective of human discourse," Σημειωτκή – *Sign Systems Studies* 2: 341-364 <http://www.ceeol.com/aspx/getdocument.aspx?id=37f0a98fa71241c790a6a1d1af38db9c>.

[16] Dappert, Angela, and Adam Farquhar (2009), "Significance is in the eye of the beholder," *13th European Conference on Research and Advanced Technology for Digital Libraries (ECDL '09)*, Corfu, September 27-October 2; also published as *Lecture Notes in Computer Science* 5714, ed. Maristella Agosti, José Borbinha, Sarantos Kapidakis, Christos Papatheodorou, and Giannis Tsakonas (Berlin: Springer-Verlag): 297-308 <http://www.planets-project.eu/docs/papers/Dappert_SignificantCharacteristics_ECDL2009.pdf>.

[17] Dewey, John (1910), *How We Think* (Boston: D. C. Heath).

[18] Doerr, Martin, and Yannis Tzitzikas (2012), *Information Carriers and Identification of Information Objects: An Ontological Approach* <http://arxiv.org/abs/1201.0385>.

[19] Eco, Umberto (1979), *A Theory of Semiotics* (Bloomington: Indiana University Press).

[20] Fedora Commons (2011), *Content Model Architecture* <https://wiki.duraspace.org/display/FEDORA37/Content+Model+Architecture>.

[21] Fedora Commons (2012), *Fedora Digital Object Model* <https://wiki.duraspace.org/display/FEDORA37/Fedora+Digital+Object+Model>.

[22] Flouris, Giorgos, and Carlo Maghini (2007), "Terminology and wish list for a formal theory of preservation," *PV 2007 – Ensuring Long-Term Preservation and Value Adding to Scientific and Technical Data*, Munich, October 9-11 <http://users.ics.forth.gr/~fgeo/Publications/PV07.pdf>.

[23] Hedstrom, Margaret, and Christopher Lee (2002), "Significant properties of digital objects: Definitions, applications, implications," *Proceedings of the DLM-Forum 2002: @ccess and Preservation of Electronic Information: Best Practices and Solutions*, Barcelona, 6-8 May (Luxembourg: Office for Official Publications of the European Communities): 218-223 <http://ils.unc.edu/callee/sigprops_dlm2002.pdf>.

[24] Heslop, Helen, Simon Davis, and Andrew Wilson (2002), *An Approach to the Preservation of Digital Records* <http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm16-47161.pdf>.

[25] IFLA Study Group on the Functional Requirements for Bibliographic Records (1998), *Functional Requirements for Bibliographic Records: Final Report*, UBCIM publications – New series volume 19 (München: K. G. Saur) <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>.

[26] Hofstadter, Douglas (1979), *Gödel, Escher, Bach: An Eternal Golden Braid* (New York: Basic).

[27] ISO 14721 (2012), *Space data and information transfer systems – Open archival information system (OAIS) – Reference model* <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=57284>.

[28] Jakobson, Roman (1960), "Closing statements: Linguistics and poetics," *Style in Language*, ed. Thomas Sebeok (Cambridge: MIT Press; New York: Wiley): 350-377.

[29] Kahn, Robert, and Robert Wilensky (1995), *A Framework for Distributed Digital Object Services*, CNRI technical report cnri.dlib/tn95-01 <http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>; also published as "A framework for distributed digital object services," *International Journal on Digital Libraries* 6:2 (April 2006): 115-123 <http://dx.doi.org/10.1007/s00799-005-0128-x>.

[30] Kant, Immanuel (1781), *Critik der reinen Vemunft* [*Critique of Pure Reason*], tr. J. M. D. Meiklejohn <http://www.gutenberg.org/files/4280/4280-h/4280-h.htm>.

[31] Library of Congress (2013), *NDSA Levels of Preservation*, version 1 <http://www.digitalpreservation.gov/ndsa/activities/levels.html>.

[32] Mackenzie, C. Matthew, Ken Laskey, Francis McCabe, Peter F. Brown, and Rebekah Metz (2006), "What is a reference model," *Reference Model for Service Oriented Architecture 1.0*, OASIS standard <http://docs.oasis-open.org/soa-rm/v1.0/soa-rm.pdf>.

[33] Mois, M., C.-P. Klas, and M. L. Hemmje (2009), "Digital preservation as communication with the future," *16th International Conference on Digital Signal Processing*, Santorini, July 5-7 <http://dx.doi.org/10.1109/ICDSP.2009.5201104>.

[34] Moore, Reagan (2008), "Towards a theory of digital preservation," *International Journal of Digital Curation* 1:3 (June): 63-75 <http://dx.doi.org/10.2218/ijdc.v3i1.42>.

[35] Morris, Charles (1946), *Signs, Language, and Behavior* (New York: Prentice-Hall).

[36] Morrissey, Sheila M. (2014), "'How can we know the dancer from the dance?' Intention and the preservation of digital objects," *New Review of Information Networking* 19.1: 1-15 <http://dx.doi.org/10.1080/13614576.2014.883935>.

[37] Nöth, Winfried (1990), *Handbook of Semiotics* (Bloomington: Indiana University Press).

[38] Peirce, Charles Sanders (1932), *Collected Papers of Charles Sanders Peirce, Volumes I and II: Principles of Philosophy and Elements of Logic*, ed. Charles Hartshorne and Paul Weiss (Cambridge: Harvard University Press/Belknap Press).

[39] Philips, Megan, Jefferson Bailey, Andrea Goethals, and Trevor Owens (2013), *The NDSA Levels of Digital Preservation: An Explanation and Uses* <http://www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA_Levels_Archiving_2013.pdf>.

[40] PREMIS Editorial Committee (2012) *PREMIS Data Dictionary for Preservation Metadata*, version 2.2 <http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>.

[41] Ranganathan, Shiyali Ramamrita (1931), *The Five Laws of Library Science* (Madras).

[42] Reynolds, Paul Davidson (1971), *A Primer in Theory Construction* (New York: Bobbs-Merrill).

[43] Rosenthal, David S. H. (2014), "Improving the odds of preservation," *CNI Fall 2014 Membership Meeting*, Washington, December 8-9 <http://www.cni.org/topics/digital-preservation/improving-the-odds-of-preservation/>.

[44] Ross, Seamus (2007), "Digital preservation, archival science, and methodological foundations for digital libraries," *11th European Conference on Digital Objects (ECDL '07)*, Budapest, September 16-21 <http://www.ecdl2007.org/Keynote_ECDL2007_SROSS.pdf>.

[45] Rusbridge, Chris (2008), "'Digital preservation' term considered harmful?" *Digital Curation Blog* (July 29) <http://digitalcuration.blogspot.com/2008/07/digital-preservation-term-considered.html>.

[46] Saussure, Ferdinand de (1916), *Cours de linguistique générale* [*Course in general linguistics*], ed. Charles Bally and Albert Sechehaye (1986; La Salle: Open Court).

[47] Schramm, Wilber (1954), "How communication works," *The Process and Effects of Communication*, ed. Wilber Schramm (Urbana: University of Illinois Press), pp. 3-26.

[48] Shannon, Claude (1948), "A mathematical theory of communication," *Bell System Technical Journal* 27 (July and October): 379-423 and 623-656 <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>.

[49] Sierman, Barbara (2014), "The SCAPRE policy framework, maturity levels, and the need for realistic preservation policy," *iPRES 2014*, *The 11th International Conference Digital Preservation*, Melbourne, October 6-10.

[50] Stamper, Rudolph (1973), *Information in Business and Administrative Systems* (New York: Wiley).

[51] Steels, Luc (2008), "The symbol grounding problem has been solved. So what's next?" *Symbols and Embodiment: Debates on Meaning and Cognition*, Manual de Vega, Arthur M. Glenberg, and Arthur C. Graesser, eds. (Oxford University Press) <http://homepage.univie.ac.at/nicole.rossmanith/concepts/papers/steels2008symbol.pdf>.

[52] Upward, Frank (1996), "Structuring the records continuum – Part one: Postcustodial principles and properties," *Archives and Manuscripts* 24:2 (November): 268-285 <http://www.infotech.monash.edu/au/research/groups/rcrg/publications/recordscontinuum-fupp1.html>.

[53] van Diessen, Raymond J. (2002), *Preservation Requirements in a Deposit System*, IBM/KB long-term preservation study report series 3 <http://www.kb.nl/sites/default/files/docs/3-preservation.pdf>.

[54] Vermaaten, Sally, Brian Lavoie, and Priscilla Caplan (2012), "Identifying threats to successful digital preservation: The SPOT model for risk assessment," *D-Lib Magazine* 19:9/10 (September/October) <http://dx.doi.org/10.1045/september2012-vermaaten>.

[55] Vitruvius Pollio, Marcus (1624), *De Architectura* [*Elements of Architecture*], trans. Henry Wotton <http://penelope.uchicago.edu/Thayer/E/Roman/Texts/Vitruvius>.

[56] Walther, Joseph, and Lisa Tidwell (1995), "Nonverbal cues in computer-mediated communication, and the effect of chronemics on relational communication," *Journal of Organization Computing and Electronic Commerce* 5:4 (November): 355-378 <http://dx.doi.org/10.1080/10919399509540258>.

[57] Wickett, Karen, Simone Sacchi, David Durbin, and Allen Renear (2012), "Identifying content and levels of representation in scientific data," *ASIS&T 75th Annual Meeting*, Baltimore, October 26-31 <http://asis.org/asist2012/proceedings/Submissions/199.pdf>.

[58] Wimsatt, William, and Monroe Beardsley (1946), "The intentional fallacy," *Sewanee Review* 54: 468-488.