A Foundational Framework for Digital Curation: The Sept Domain Model

Stephen Abrams
California Digital Library
University of California
Oakland, CA 94612, US
+1 510-987-0370
Stephen.Abrams@ucop.edu

ABSTRACT

Digital curation is a complex of actors, policies, practices, and technologies enabling successful consumer engagement with authentic content of interest across space and time. While digital curation is a rapidly maturing field, it still lacks a convincing unified theoretical foundation. A recent internal evaluation by the University of California Curation Center (UC3) of its programmatic activities led quickly to seemingly simple, yet deceptively difficult-to-answer questions. Too many fundamental terms of curation practice remain overloaded and under-formalized, perhaps none more so than "digital object." To address these concerns, UC3 is developing a new model for conceptualizing the curation domain. While drawing freely from many significant prior efforts, the UC3 Sept model also assumes that digital curation is an inherently semiotic activity. Consequently, the model considers curated content with respect to six characteristic dimensions: semantics, syntactics, empirics, pragmatics, diplomatics, and dynamics, which refer respectively to content's underlying abstract meaning or emotional affect, symbolic encoding structures, physical representations, realizing behaviors, evidential authenticity and reliability, and evolution through time. Correspondingly, the model defines an object typology of increasing consumer utility and value: blobs, artifacts, exemplars, products, assets, records, and heirlooms, which are respectively existential, intentional, purposeful, interpretable, useful, trustworthy, and resilient digital objects. Content engagement is modeled in terms of creator, owner, curator, and consumer roles acting within a continuum of concerns for catalyzing, organizing, and pluralizing curated content. Content policy and strategy are modeled in terms of seven high-level imperatives: predilect, collect, protect, introspect, project, connect, and reflect. A consistent, comprehensive, and conceptually parsimonious domain model is important for planning, performing, and evaluating programmatic activities in a rigorous and systematic rather than ad hoc or idiosyncratic manner. The UC3 Sept model can be used to make precise yet concise statements regarding curation intentions, activities, and results.

General Terms

Frameworks for digital preservation.

Keywords

Digital curation, digital preservation, domain model, semiotics, continuum, policy, strategy.

1. INTRODUCTION

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

A given unit of 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 *successful* if the content can be feasibly exploited for use and that use is beneficial for some desired purpose, ideally at a time and place and in a manner of the consumer's choosing. Feasibility of use depends upon intellectual and technical considerations regarding production and management, for example, selection, acquisition, arrangement, integrity, permission, visibility, etc., while the benefit of use is conditioned by individualistic purpose. It is possible that this purpose may be fulfilled only at some considerable spatio-temporal distance from the point of the content's creation; regardless, the consumer's purpose, and derived benefit, is not necessarily constrained to conform to the original intention of the content's creator, owner, or steward. Rather, every engagement is uniquely situated with respect to the context of the content's production, its curatorial framing, and its consumer's collateral experience, expertise, and expectation. Although this context is ultimately subjective, it may nevertheless be commonly held by other consumers participating in the same domains of discourse.

The curation attributes of enablement, success, engagement, authenticity, and interest are a contemporary restatement of traditional content stewardship concerns as articulated, for example, by Ranganathan's "laws" of library science [29]. The first law, "Books are for use," shorn of its biblio-centricity, is fundamentally concerned with *utility*, that is, the use for purpose underlying any successful engagement with a message-bearing object. 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. 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 content's ease of mutability. Mutability of a different sort is 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 convenience, or more generally, *service*, and corresponds to the imperative of curating agents providing their customers with tools and services that effectively and efficiently meet their intellectual, behavioral, and technical expectations. Underlying all of these concerns is the notion that curation encompasses both preservation and use [42] [33], which are

complementary rather than disparate activities: preservation ensuring use *over* time while use is dependent upon preservation *up until a point* in time.

Curation outcomes naturally lie along a spectrum of possible results largely dependent upon the degree to which appropriate human, organizational, and technical resources can be applied. 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 of content based on its assessed value and institutional priority.

2. MODELING THE CURATION DOMAIN

Curation decisions should be made with respect to an underlying theory or conceptual domain model. A domain model is an abstraction of fundamental expressive and explanatory principles sophisticated enough to explicate past events and conditions and anticipate the consequences and efficacy of future decisions and actions; in other words, it should be both descriptive and predictive [30]. It is useful to build up such a model incrementally from first principles in order to ensure comprehensive scope, self-consistency, and conceptual parsimony. It is important, however, to keep in mind that all models are at best idealized representations of nominal domain concepts. The simplifying assumptions and abstractions inherent to any modeling effort may be at times incommensurate with pertinent real world detail and any actual curation entity or condition may not fully conform to model formalisms.

When the University of California Curation Center (UC3) first started a comprehensive internal review of its curation activities to evaluate their efficacy and set future priorities, it did so in the context of many descriptive and prescriptive frameworks familiar to the digital curation and preservation communities, for example, the ISO 14721 OAIS reference model, PREMIS, TRAC, etc. In working with these models, however, UC3 staff soon found themselves asking a number of seemingly simple, yet deceptively difficult-to-answer questions. What *exactly* is a "digital object"? (A bitstream? A file? A package? A dissemination?) What *specifically* is meant by "preservation" of an object? (A concern for the integrity of bits? Of context? Of performance? Of understanding?) None of the preexisting frameworks provided fully sufficient answers. In addition to definitional ambiguity, it was not immediately apparent how – or indeed whether – the conceptual models underlying these disparate efforts cohered into a unified and inclusive picture of the curation domain. A comprehensive reference model is important in ensuring that programmatic curation activities are planned, performed, and evaluated in a rigorous and systematic rather than ad hoc or idiosyncratic fashion. To address these concerns, UC3 has developed a new approach towards conceptualizing the curation domain that draws freely from past efforts, but also incorporates applicable concepts from other relevant fields such as information science, cognitive psychology,

and semiotic theory. The UC3 Sept model affords a useful conceptual map, analytical framework, and descriptive vocabulary applicable to the full range of curation activities [41].

2.1 Curation semiotics

The ultimate goal of curation is to facilitate the effective "delivery" of content to human consumers across barriers imposed by space and time. (Even in cases of intermediating technical systems, ultimate agency always resides in a human actor [10].) In psycho-physiological terms, an act of content consumption occurs when:

1. An abstract unit of content is …
2. Realized by physical stimuli, which are …
3. Perceived by a sense modality, …
4. Interpreted in the specific subjective context of the consumer, and ultimately …
5. Experienced as cognitive meaning or psychological affect.

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

Semiotics is the study of signs and systems of signification, that is, things that *carry* messages and the ways in which those messages are represented and communicated [21] [25]. A sign is something that "stands in" for something else, in some manner, for someone [26]. In other words, it is a triadic relation between an external referent, its representation, and its effect on the consumer, which is a new mental state or reformulation of the referent stimulated by its representation. This cognitive or emotional state always arises in the subjective contextual ground of the consumer's collateral experience independent of the sign itself [5]. No unit of content is inherently significant; it gains significance for a given consuming agent only "in a context relevant to some purpose or goal" [12].

2.1.1 Roles

The consumer role is defined in the generic sense of an actor who derives some benefit from the direct use of, or indirect reference to, curated content. Direct exploitation may be consumptive (for example, passive reading, watching, listening to, etc.), generative (creating something new), or manipulative (adding to, modifying, or deleting from something extant). Indirect benefit, on the other hand, may be derived merely from the existence of content independent of direct use. The retention of certain legal materials, for example, confers tangible value to agents subject to relevant statutory or regulatory obligations or those with a legal interest in the materials' subject matter. The other fundamental curation roles are content creator, content owner, and content curator, corresponding to agents exercising creative, proprietorial, and delegated stewardship responsibilities, respectively. Any or all of these roles may be held by a given individual or corporate actor at various times and varying organizational and operational contexts.

2.1.2 Analytical concerns

For purposes of analysis, it is useful to consider digital content in terms of six characteristic dimensions: semantics, syntactics, empirics, pragmatics, diplomatics, and dynamics:

1. Semantics is concerned with the relationships between content and its underlying abstract meaning or affect;
2. Syntactics, with the relationships between content and its symbolic expressions;
3. Empirics, with the relationships between content and its physical representations [38];
4. Pragmatics, with the relationships between content and its consumers, that is, those concerning realizing

behaviors [24];

5. Diplomatics, with the relationships between content and the factual authenticity and reliability of its expression, representation, management, and transmission [32]; and

6. Dynamics, with the relationships between various states of content as it persists and evolves across space and time [9] [16].

(The term "diplomatics" is used here as a convenient generic label for a complex of concerns regarding trustworthiness rather than the more specific sense of use common to archival practice.)

These analytic dimensions correspond to longstanding stewardship concerns with content's abstract meaning, symbolic inner structure and outer form, physical carrier, behavioral experience, archival authenticity, and spatio-temporal persistence. They also give rise to the "Sept" model name, which was suggested by the approximate phonetic pronunciation of the SSEPDD dimensional acronym. "Sept" is also a genealogical term referring to a subgroup of an extended clan or family, appropriate for a model concerned with delineating nuanced distinctions within digital objects.

2.2 Object modeling

Digital objects are encapsulations of information for purposes of communication. Before devising Sept, UC3 reviewed a number of prior models for objects and the more general notion of communicable information, including the sender/receiver model (that is, Shannon and Weaver [36] as extended by Schram [35] and Berlo [6]); Buckland's information trichotomy [7]; Kahn and Wilensky [20], FRBR [18], the NAA performance model [17], OAIS [19], PREMIS [28], the Basic Representation Model (BRM) [43], and the Information Carrying Ontology (ICO) [14]. The component ontological subdivisions defined by these models can be approximately aligned against one another in a tabular fashion as shown in Table 1.Two pertinent facts emerge from this exercise: first, the number of rows in the table indicates the overall fineness of granularity with which these models have usefully decomposed the concept of an information object; and second, none of the prior efforts completely addresses the full gamut of ontological concerns at the finest decompositional level. The Sept model is intended to unambiguously defining of all ontological granules in a single coherent model, clarifying what an object is and what it is not.

Table 1. Information modeling crosswalk

Sender / receiver (1960)	Buckland (1991)	Kahn-Wilensky (1995)	FRBR (1998)	NAA (2002)	OAIS (2003)	PREMIS (2005)	BRM (2012)	ICO (2012)	UC3 Sept (2015)
source	info-as-knowledge		work	essence	information object	intellectual entity	propositional content	intellectual entity	message
encoding	info-as-thing	data	expression	source	data object / digital object	bitstream / filestream	symbol structure	symbol structure	structure
			manifestation			file / representation			form
			item		bits		patterned matter/energy	information carrier	carrier
frame-of-reference		key-metadata			representation information		auxiliary information		annotation
channel	info-as-process			process				projection	behavior
signal				performance				sensory impression	stimuli
context					knowledge base				ground
decoding									interpretation
effect	info-as-knowledge		work	essence	information object	intellectual entity	propositional content	intellectual entity	experience

2.2.1 Message vs. meaning

An object is a means by which its creator intends to communicate with a consumer. However, while an object can convey the creator's *message* – the numbers, words, images, sounds, etc. that constitute its information content – the *meaning* ascribed to that message is not actually carried by the object itself. Rather, the consumer's experience of cognitive meaning or emotional affect is an emergent epistemic effect of the consumptive act. An object mediating that act is a reflection of a particular mental state of its creator and is intended to induce a corresponding state on behalf of its consumer. However, since the consumer's mental formation of meaning arises through a contextually-grounded interpretation of

the object's message, the creator's intention may never be fully realized [44]. While the potential for discordant interpretation may be minimal regarding the communication of propositional content, that is, content pertaining to objectively factual truth claims, individualistic responses are accepted and often even highly desirable outcomes for engagement with creative expressions.

2.2.2 Ontological components

In semiotic-theoretic terms, an act of object-mediated communication occurs when an expressible message is encoded into an object susceptible to contextualized decodings, resulting in subjectively experienced cognitive meanings or emotional affects (see Figure 1). In other words, an object reifies an abstract

expressible thought, relative to some contextual frame of reference, into a consumable embodied thought, a critical distinction long established in the semiotic field, viz., *parole* vs. *langue*, or signifier vs. signified [25], as well as in library and information science, viz.,

work vs. document [39]. Following from this, the major ontological components of a digital object are its message, encoding of structure and form, carrier, behavior, and annotation, reflecting the chain of content reification (see Figure 2).
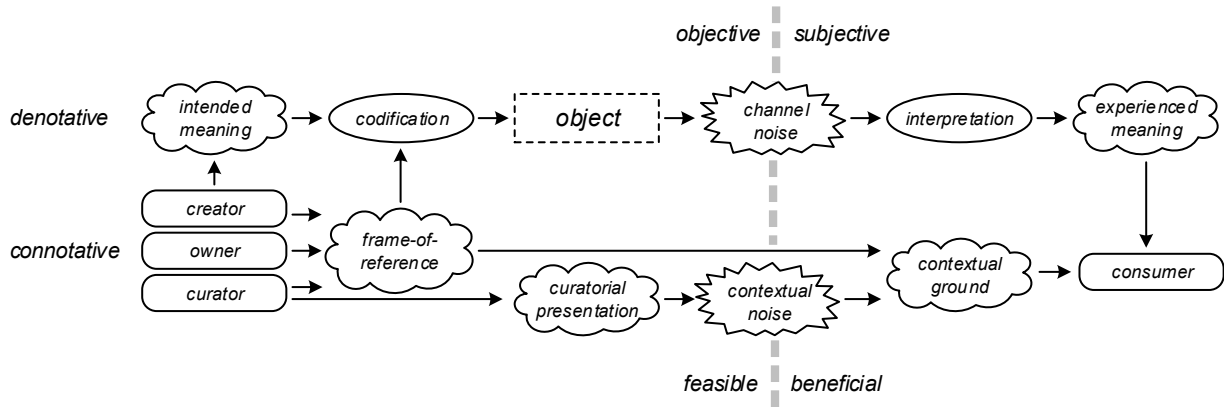

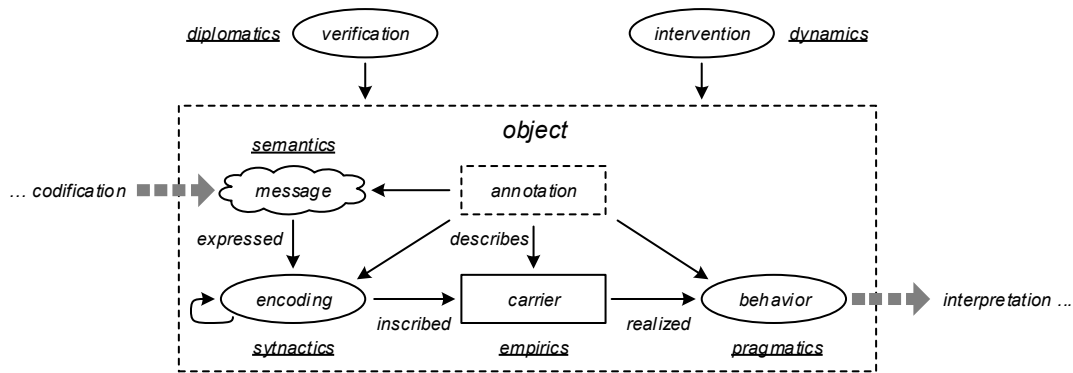**Figure 1. Object-mediated communication**

Figure 2. Object components

An object's message constitutes its semantic aspect, that is, the abstract information content it is intended to convey. This content is *expressed* through encodings into abstract symbol structures constituting the object's syntactic aspect [42] [14]. A given object may be distinguished by multiple hierarchically-nested encodings. These encodings can be distinguished between those concerned with the object's inner structure and outer morphological form. An object's symbolic expression is given tangible representation by being *inscribed* as a pattern of matter or energy on a physical carrier that constitutes the object's empiric aspect. This physical representation is made available for perception and interpretation by a consumer by being *realized* through behaviors that render the underlying information content in a human-sensible manner, constituting the object's pragmatic aspect.

The primary attribute of an object's inner structure is its format or type, which specifies the conventions of the object's symbolic expression and is the interface with its semantics [1]. The primary attribute of object morphology is identity. The identity of an object, like that of a linguistic sign, serves three purposes [13]:

1. As a *fence*, distinguishing and demarcating a particular object from all other potential objects;
2. As a *label*, facilitating unambiguous common reference to a singular object; and

3. As a *vehicle*, providing an actionable means for interacting with the object for some teleological purpose.

Morphological form also implicates the interface between the object and its empirics, that is, the encompassing computational environment necessary to support the object's visibility and dereferencing, for example, encoding details attendant to a file system, run-time environment, or network infrastructure depending upon whether the object is at rest or in motion. Without an assertion of identity, there is no effective way to establish or retrieve an object as the focus of curation scrutiny; similarly, without format typing, there is no effective means of interpreting and exploiting the object's message.

The attributes of identity and type are instances of annotations, propositional statements declaring specific characteristic values for significant object properties [12]. As these descriptive properties are fundamental to the successful interpretation and exploitation of an information object, they are a type of OAIS representation information [19]. (Representation information is also concerned with instrumental capabilities, for example, a viewer for a particular type; these are equivalent to Sept's pragmatic behaviors.) While identity and type are fundamentally *necessary* annotative properties, by themselves they may not be fully *sufficient* to ensure successful engagement, which may be dependent upon additional

higher-order semantic and pragmatic properties [2]. Annotations provide the means to assert the perspectives or frames-of-reference of an object's creator, owner, and curatorial manager, and the interpretive experience of its consumer. The relationship between an annotation and its referent foreshadows that between the consumer's ground and interpretation: annotations contributing to the content's *objective* context and collectively informing the interpretive ground of the consumer's *subjective* experience.

2.2.3 Contextual ground

Traditional semiotic analysis presupposes two primary actors in the communicative act: creators and consumers. When dealing with curation of digital content, however, a third curatorial actor often intermediates between creator and consumer at the behest of a fourth, the content's owner. Content is traditionally collected, managed, and presented for use by a curatorial agent as part of larger aggregations based on explicit collection development policy, thematic unity, or administrative convenience. The contextual perspectives of the curatorial manager inevitably leave traces in a consumer's interpretive experience, just as a creator and owner's conceptual frames-of-reference inform the intention underlying content production. Thus, an object is inherently situated within a dynamic network of explicit and implicit denotative, connotative, and metaphorical associations by which it accumulates additional nuanced meanings or affects through the circumstances of its production, membership in curatorial aggregations, and under the imprimatur of its stewarding manager. Denotation refers to the overt commonplace meaning of an expression; connotation, to the indirect inferential meaning; and metaphor, to the allusive meaning [11]. These associations can take several forms:

1. Reputational assessments of individual and corporate content creators, owners, and curatorial stewards arising from a consumer's prior experience and professional judgment;
2. Intentions attendant to content production and ownership as expressed in collateral annotations;
3. Intentions attendant to content management as expressed through curatorial assessment, selection, arrangement, aggregation, and annotation;
4. Tangible relationships directly representable in content objects or object management systems, such as typed structural and semantic relationships between separate but dependent objects, and object aggregations and their subsidiary members; and
5. The tacit understanding acquired through experience or, in OAIS terms, as part of the knowledge base of a designated community that a consumer brings to the interpretive act [19].

All of these factors contribute to, but do not fully determine, the consumer's interpretive ground and subjective experience. The inherently recursive nature of these associational networks, in which every referent potentially can be the source of further references, is reminiscent of OAIS representation networks.

2.3 Object typology

Not every object will consist of the full complement of components. Thus, it is useful to distinguish between seven characteristic object types: blobs, artifacts, exemplars, products, assets, records, and heirlooms (see Table 2), which constitute a typology of increasingly specific definition and functional utility and value.

Table 2. Object typology

Differentia	Blob	Artifact	Exemplar	Product	Asset	Record	Heirloom
Dimension	empirics	syntactics	syntactics	semantics	pragmatics	diplomatics	dynamics
Mode	formative	informative	informative	informative	performative	evaluative	reformative
Act	inscription	identification	characterization	description	realization	verification	intervention
Concern	media	(outer) encoding	(inner) encoding	meaning / affect	experience	authenticity	persistence
Abstraction	carrier	form	structure	message	behavior	evidence	action
Quality	existential	intentional	purposeful	interpretable	useful	trustworthy	resilient
Value	nascent	incipient	potential	theoretical	practical	assured	enduring
Annotation	provenancial / administrative / permissive	morphological / relational / associational	structural	intellectual	instrumental	provenancial	provenancial

1. A blob is an *existential* object resulting from a formative act of inscription that produces tangible bits on an otherwise undifferentiated digital carrier, whether storage media or communication channel. Being opaque in all respects, nothing further can be known or inferred about a blob other than the fact of its existence. Thus, its value is nascent. Consider, for example, the bits …00000000110001101010010100… found somewhere on a carrier, which by themselves convey no recognizable, let alone useful, information.
2. An artifact is an *intentional* object resulting from an

informative act of identification that demarcates a particular sequence of bits fixed in digital space-time. One can infer that an artifact was deliberately created, even if the purpose underlying the creation remains undisclosed. The essential properties of an artifact are its identity and symbolic encoding of outward-facing morphology. 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 message. Thus, an artifact's value is incipient. Consider, for example, a named file with specific size,

timestamps, permissions, and MD5 digest, but absent any knowledge of its content's expression.

3. An exemplar is a *purposeful* object resulting from an informative act of characterization that documents the symbolic encodings of its internal structural expression. (The term "exemplar" is used in here in its non-qualitative sense of a general pattern or template without individuating characteristics.) The essential properties of an exemplar are its type or format and any further attributes entailed by that format. While these provide details of the exemplar's means of expression, its underlying message is still semantically opaque. Thus, an exemplar's value is potential. Consider, for example, a JPEG 2000-formatted image with three 8-bit components representing sRGB color samples, with 1024x1024 tiles, 64x64 code blocks, six decomposition layers, 25 quality layers, and 9-7 irreversible wavelet compression, but absent any knowledge of what the image represents.

4. A product is an *interpretable* object resulting from an informative act of description that documents its underlying message in terms appropriate to a particular domain of discourse. In and of itself, however, a product doesn't afford any practical means to experience or exploit that message. Thus, a product's value is theoretical. Consider, for example, 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 realizing behaviors.

5. An asset is a *useful* object resulting from a performative act of realization that exposes the product's message as stimuli apprehensible to human sensory modalities [5] [17]. Thus, an asset's value is practical: it can be directly experienced and exploited towards some useful purpose. Consider, for example, a consumer's experience engaging with the authentic Lake Merritt image in a colorimetric image processing environment supporting dynamic zooming, panning, cropping, annotation, etc., but absent any consideration of spatial or temporal extension.

6. A record is a *trustworthy* object resulting from an evaluative act of verification. The essential properties of a record are those important to considerations regarding the presumption, verification, and maintenance of authenticity and reliability [15]. Being trustworthy, a record's value is assured. Consider, for example, the Lake Merritt image that has been evaluated and determined to be what it purports to be, so that it can be accepted with confidence.

7. An heirloom is a *resilient* object resulting from a reformative act of proactive or reactive intervention that ensures the continuing viability and usability of the asset across space and time. Thus, to the extent to which those interventions are successful, an heirloom's value is enduring. Consider, for example, a consumer's future engagement experience with the Lake Merritt image.

The encodings underlying artifacts and exemplars may be hierarchically nested, for example, an artifact that is a file in a folder on a disk volume, or an exemplar that is a PCM sample stream inside a QuickTime multimedia wrapper inside of a Zip container.

The sequence of object types from blobs to heirlooms provides increasing functional utility and value, but the typology does not imply a strictly sequential inheritance hierarchy. While in practice many digital objects will have valid ontological identities across contiguous typological classes – for example, a product with known semantics, encoded in a known format (and thus, also an exemplar) and in well-characterized file (and thus, an artifact), inscribed on some tangible media (and thus, a blob) – this is not a necessary condition of the Sept model. Any higher-order type can effectively subclass directly from any inferior type. It is possible, for example, for product semantics to be known about an article whose inner encoding remains syntactically opaque. For example, consider an object about which the statement "This file is an image (of unknown format) of Lake Merritt" can be made. While one might have cause to question the accuracy of the assertion, it is nevertheless a valid case of a product being an artifact but not an exemplar. Similarly, it is possible for an asset to be an exemplar but not a product ("This JPEG image (of unknown subject) is viewable in that JPEG viewer"), a record to be a product but not an asset ("This image (with no format-specific viewer available) really *is* of Lake Merritt and *was* produced by the Lake Merritt Breakfast Club Foundation"); and an heirloom to be an asset but not a record ("This persistent viewable JPEG image *may* be of Lake Merritt").

2.3.1 Resilience

Resiliency ensures that an heirloom can be used for successful "communication with the future" [23] [22]. In information theory, factors that impede communication are considered noise [36]. In planning for effective interventions to ensure resiliency, the information-theoretic sender/receiver communication model distinguishes between *channel* and *contextual* noise: the former degrades the integrity of the signal, that is, the object carrying the encoded message, while the latter distorts the interpretive context of the object's message – for example, a conceptual misalignment between 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 object, for example, mirroring, parity, checksums, erasure codes, etc. A strategy for minimizing contextual noise is to facilitate the most effective means for the creator, owner, and curatorial frames-of-reference to inform fully the contextual ground of the consumer; in other words, to ensure that the consumer can properly recover productive, proprietorial, and curatorial intentions. Descriptive annotations are included as a fundamental component of a digital object in order to facilitate this very process. However, since this strategy implies communication of the 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.

2.3.2 Annotation

Annotations are defined in terms of nine high-level categories: provenancial, administrative, relational, associational, 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 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 discourse; and
9. Instrumental annotations describe behaviors applicable to the content.

Table 2 indicates the earliest stage in the typological progression at which those particular annotation categories are relevant. For example, a blob has provenancial properties independent of and prior to any artifactual concerns (for example, carrier *A* was received from agent *B* at time *C*, etc.), an artifact has morphological properties independent of and prior to any exemplar-level concerns (file *X* of size *Y* and modification date *Z*, etc.), and so on.

2.4 Content engagement

Engagement with digital content is modeled in terms of four classes of actors and the lifecycle activities in which they participate [3] [8]. Content creators generate or acquire digital content and exercise *originating* intellectual and instrumental control and responsibility for the circumstances of that creation or acquisition; content owners exercise *ultimate* legal, financial, and permissive control and responsibility for its ongoing stewardship; content curators steward managed content and exercise *delegated* administrative, technical, and instrumental control and responsibility; and content consumers directly exploit or indirectly benefit from managed content for some individualistic purpose. The creator, curator, and consumer roles have a general correspondence to the producer, management, and consumer entities in the OAIS reference model [19]. The Sept consumer role, however, is more inclusive than its OAIS counterpart, encompassing any agent gaining some benefit from curated content through either direct *or* indirect means; while the Sept owner role and its concerns of proprietorial rights and obligations are not directly represented in the OAIS model. A given individual or corporate actor may hold these roles singly or a varying combinations at different points of time and in different organizational and operational contexts.

It is more useful to speak of the concerns of these roles in terms of an activity *continuum* rather than a lifecycle, as the latter implies a linear progression through clearly demarcated and distinguishable stages. In distinction, a continuum approach emphasizes the essential non-linear contiguity and overlapping interdependence of many curation activities and concerns [40]. Thus, it is more appropriate to group modes of engagement by thematic loci of concerns within a permeable continuum characterized by first-order catalyzation, concerned with creating, acquiring, or otherwise establishing resources of curation focus; second-order organization, concerned with codifying and imposing illuminating structure upon those resources; and third-order pluralization, concerned with expanding the reach and consequence of those resources. These continuum characteristics are based on the information continuum model (ICM) [34] although Sept's notion of catalyzation conflates the ICM's distinct creation and capture dimensions into a single category for purposes of conceptual parsimony.

While the thematic loci and continuum characteristics may seem synonymous – for example, production being equivalent to catalyze, etc. – they are actually orthogonal concerns: as indicated in Table 3, activities within each locus can be categorized by goals and intents spanning all three characteristic categories. Similarly, although terminological similarity implies a reductive association between roles and loci, for example, creators and production, etc., these are also orthogonal concerns, with each locus encompassing activities spanning each role. (Unfortunately, it is difficult to provide an intuitive depiction of the mutually-interdependent relationships of the three disparate dimensions of role, locus, and characteristic in a simple tabular form.) A comprehensive curation program will work towards promulgation of policies, strategies, and plans, and implementations of systems, services, operational procedures, and stakeholder guidance for all major continuum activities.

Table 3. Engagement continuum

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

2.5 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 intentions for fulfilling or enforcing promulgated policies that can be implemented by concrete plans and activities [4] [37]. Curation policies, strategies, plans, and activities are modeled within Sept in terms of one preparatory and six implementation imperatives:

1. *Predilect*: decide what you intend;
2. *Collect*: obtain or effectuate what you decide;
3. *Protect*: preserve or sustain what you obtain;
4. *Introspect*: know what you protect;
5. *Project*: offer what you know; and
6. *Connect*: provide what you offer.
7. *Reflect*: (re)assess what you did.

While these imperatives are relevant to all aspects of the curation domain, for example, technical infrastructure, operational procedures, staffing, etc., they have the most obvious applicability to content. There is a general inheritance of relevant considerations across the imperative progression. The range of activities underlying these imperatives span the speculative and exploratory

(that is, considerations of what *could* be), analytical and normative (what *should* be), operational (what *is*), and obligatory (what *must* be). In general, these activities should be proactive whenever possible, and reactive whenever necessary.

The foundational imperative for subsequent curation activities is *collection*, that is, bringing content into an appropriate stewardship environment under the control of a responsible curatorial manager with rights and obligations delegated from the content's owner. While it is *possible* that collected content may not be fully susceptible to successful curation outcomes, it is almost *certain* that uncollected content will be subject to curation failure with regard to current or future viability and availability [31]. The baseline level of curation assurance that can be realistically asserted by a responsible curating agent will generally be either as a blob or artifact, depending upon whether the content was collected as (undifferentiated) media or (opaque) files. Increasingly high-order

outcomes may be possible if the content meets the incrementally more stringent criteria for exemplars, products, assets, records, and heirlooms.

Each imperative can be applied to every level of the typological hierarchy. While the resulting matrix (see Table 4) is suggestive of the NDSA levels of preservation [27], the typological progression plays a different role than the NDSA levels as it is defines increasing levels of general *utility* rather than specific *assurance*. However, concerns of assurance are encapsulated within the protect imperative. Thus, content utility and assurance both increase through the effective provisioning and implementation of progressive levels of environmental, administrative, technical, bibliographic, archival, access, and change control. Similarly, utility and assurance both increase through progressive levels of forensic, morphological, structural, intellectual, archival, and behavioral characterization arising from the introspect imperative.

Table 4. Policies and strategies

Imperative	Blob	Artifact	Exemplar	Product	Asset	Record	Heirloom
Predilect	service level agreement	disaster recovery / business continuity	format action plans	collection development policy	outreach and training	evidentiary standards	sustainability / succession planning
Collect	submission	packaging	normalization / canonicalization	discovery, workflow / tool integration	code / workflow repositories, aggregation	provenance	preservation planning tools
Protect	environmental control, media refresh, redundancy	administrative control, malware detection, fixity	technical control, migration	bibliographic control	access control, emulation	archival control	change control, preservation watch
Introspect	forensic characterization	morphological characterization, PID minting	structural characterization, ontologies, format registries	intellectual characterization, entity extraction, sentiment analysis	behavioral characterization, software registry	archival characterization, master registry	provenance, annotation
Project	media inventory	file inventory, PID resolution	object index	work catalog	transcoding, syndication, discovery	documentary form	versioned change history
Connect	legacy/emulated computational environments	file delivery	local format-aware processing	local disciplinary-specific processing	search/browse, hosted tools, annotation	authenticity-dependent workflows	consortial collaboration
Reflect	scrubbing	audit	tabletop testing	policy conformance	analytics	chain of custody	failure injection

3. CONCLUSION

The digital curation field has reached a stage of maturity where it can usefully draw upon a rich body of 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 ensure 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 into a new inclusive model. One important insight of the UC3 Sept modeling effort is that engagement with digital content is an inherently semiotic activity. Thus, the Sept model was developed by approaching all aspects of the curation domain through the lens of six characteristic dimensions: semantics, syntactics, empirics, pragmatics, diplomatics, and dynamics. The model conceives of a digital object as reifying abstract content into tangible form for purposes of mediated communication between a

creator and consumer, carefully distinguishing between an object's message and meaning; the former being an objective embodiment of an expressed thought, while the latter is an emergent epistemic property arising from a subjective, contextualized reaction to the message. This leads to an object typology of progressively richer ontological basis and concomitant increasing content utility and value, consisting of blobs, artifacts, exemplars, products, assets, records, and heirlooms. Engagement with curated content is modeled by creator, owner, curator, and consumer agents and three loci of concerns for production, management, exploitation all operating within a continuum of originating, organizing, and pluralizing dimensions. Curation policies and strategies are modeled by seven imperatives: predilect, collect, protect, introspect, project, connect, and reflect.

The model components and its typology represent useful abstractions whose properties, coalescing around core conceptual centers of gravity, may be held by any particular component or

typological instantiation. The components and typology can be used to make precise yet concise assertions regarding programmatic capabilities, intentions, actions, and outcomes. 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 blobs, exemplars, and assets. Creating forensic disk images is a suitable strategy for preserving blobs (that is, media objects), independent of any artifactual morphology; fixity audit is a suitable strategy for artifacts (file objects), independent of any type characterization; migration, for products (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, a digital exemplar (that is, 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 exemplar 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 exemplar makes clear that it will continue to be a purposeful object through persistent association with pertinent inner structural encoding information. Similarly, a preserved product will remain interpretable 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 components on the objective side of the communication divide, i.e., message, encoding, carrier, annotation, and behavior (see Figure 1), on the subjective side, the consumer's future contextual ground is not susceptible to any equivalent constraint as it is contingent on the totality of that consumer's intervening lived experience. This may not be significant for propositional content consisting of purportedly-objective factual claims, but could be important for creative content.

All of the Sept model components were developed incrementally from first principles in an effort to ensure comprehensive applicability and internal consistency. The use of such a model is important for increasing confidence that programmatic planning is systematic and not ad hoc. 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 Sept 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.

4. ACKNOWLEDGMENTS

5. REFERENCES

[1] Abrams, S. 2007. File formats. In *Curation Reference Manual*, Digital Curation Centre, URL=http://www.dcc.ac.uk/resources/curation-reference-manual/completed-chapters/file-formats.

[2] APARSEN. 2014. *D11.3 report on a common vision of digital preservation: Progress to year 3*. Technical report. URL= http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2014/06/APARSEN-REP-D11_3-01-1_1_inclURN.pdf.

[3] Ball, A. 2012. *Review of Data Management Lifecycle Models*. University of Bath. URL=http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf.

[4] Becker, C., Kulovits, H., Guttenbrunner, M., Strodl, S., Rauber, A., and Hoffman, H. 2009. Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans. *Intl. J. Dig. Lib.* 10, 4 (December), 133-157. DOI=doi:10.1007/s00799-009-0057-1.

[5] Benyon-Davis, P. 2011. *Significance: Exploring the Nature of Information, Systems and Technology*. Palgrave Macmillan, New York.

[6] Berlo, D. 1960. *The Process of Communication*. Holt, Rinehart, and Winston, New York.

[7] Buckland, M. 1991. Information as thing. *J. Am. Soc. Inform. Sci.* 42, 5 (June), 351-360.

[8] CEOS Working Group on Information Systems and Services. 2012. *Data Life Cycle Models and Concepts*. URL=http://www.ceos.org/images/DSIG/Documents/Data_Lifecycle_Models_and_Concepts_v13-1.docx.

[9] Cheney, J., Lagoze, C., and Botticelli, P. 2001. Towards a theory of information preservation. In *5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL '01)* (Darmstadt, September 4-9). HDL=http://hdl.handle.net/1853/5828.

[10] Dallas, C. 2007. An agency-oriented approach to digital curation theory and practice. In *ICHIM ''07, International Cultural Heritage Informatics Meeting* (Toronto, October 24-26), 49-72. URL=http://www.archimuse.com/ichim07/papers/dallas/dallas.html.

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

[12] Dappert, A., and Farquhar, A. 2009. Significance is in the eye of the beholder. In *Proceedings of the 13th European Conference on Research and Advanced Technology for Digital Libraries* (Corfu, September 27-October 2, 2009). URL=http://www.planets-project.eu/docs/papers/Dappert_SignificantCharacteristics_ECDL2009.pdf.

[13] Dewey, J. 1910. *How We Think*. D. C. Heath, Boston.

[14] Doerr, M., and Tzitzikas, Y. 2012, *Information Carriers and Identification of Information Objects: An Ontological Approach*. Technical report. URL=http://arxiv.org/abs/1201.0385.

[15] Duranti, L., ed. 2005. *The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*. Archilab, San Miniato. URL= http://www.interpares.org/book/index.htm.

[16] Flouris, G., and Maghini, C. 2007. Terminology and wish list for a formal theory of preservation. In *PV 2007 – Ensuring the Long-Term Preservation and Value Adding to Scientific and Technical Data* (Munich, October 9-11). URL=http://www.pv2007.dlr.de/Papers/Flouris_WishListPre servation.pdf.

[17] Heslop, H., Davis, S., and Wilson, A. 2002. *An Approach to the Preservation of Digital Records*. National Archives of Australia. URL=http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm16-47161.pdf.

[18] IFLA Study Group on the Functional Requirements for Bibliographic Records. 1998. *Functional Requirements for Bibliographic Records: Final Report*. K. G. Saur, München. URL=http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2 008.pdf.

[19] ISO 14721. 2012. *Space data and information systems – Open archival information system (OAIS) – Reference model*. URL= http://public.ccsds.org/publications/archive/650x0m2.pdf.

[20] Kahn, R., and Wilensky, R. 1995. A framework for distributed digital object services. *Intl. J. Digital Libraries* 6, 2 (April), 115-123. DOI=http://dx.doi.org/10.1007/s00799-005-0128-x.

[21] Liebenau, J., and Backhouse, J. 1990. *Understanding Information: An Introduction*. Macmillan, London.

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

[23] Moore, R. 2008. Towards a theory of digital preservation. *Intl. J. Digital Curation* 3, 1 (June), 63-75. DOI=http://dx.doi.org/10.2218/ijdc.v3i1.42.

[24] Morris, C. 1946. *Signs, Language, and Behavior*. Prentice-Hall, New York.

[25] Nöth, W. 1990. *Handbook of Semiotics*. Indiana University Press, Bloomington.

[26] Peirce, C. S. 1932. *Collected Papers of Charles Sanders Peirce. Volumes I and II: Principles of Philosophy and Elements of Logic*. Harvard University Press, Cambridge.

[27] Philips, M, Bailey, J., Goethals, A., and Owens, T. 2013. *The NDSA Levels of Digital Preservation: An Explanation and Uses*. Library of Congress. URL=http://www.digitalpreservation.gov/ndsa/working_grou ps/docuemnts/NDSA_Levels_Archiving_2013.pdf.

[28] PREMIS Editorial Committee. 2012. *PREMIS Data Dictionary for Preservation Metadata*. URL=http://www.loc.gov/standards/v2/premis-2-2.pdf.

[29] Ranganathan, S. R. 1931. *The Five Laws of Library Science*. Madras Library Association.

[30] Reynolds, P. D. 1971. *A Primer in Theory Construction*. Bobbs-Merrill, New York.

[31] Rosenthal, D. S. H. 2014. Improving the odds of preservation. *CNI Fall 2014 Membership Meeting* (Washington, December 8-9). URL=http://www.cni.org/topics/digital-prservation/improving-the-odds-of-preservation.

[32] Ross, S. 2007. Digital preservation, archival science, and methodological foundations for digital libraries. In *ECDL 2007, The 11th European Conference on Digital Libraries* (Budapest, September 16-21). URL=http://www.ecdl2007.org/Keynote_ECDL2007_SROS S.pdf.

[33] Rumsey, A. S. 2010. Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information. URL=http://blueribbontaskforce.sdsc.edu/biblio/BRTF_Final _Report.pdf.

[34] Schauder, D., Johanson, G., and Stillman, L. 2005. Sustaining a community network: The information continuum, and the case of VICNET. *J. Community Informatics* 1, 2. URL=http://ci-journal.net/index.php/ciej/article/view/239/203.

[35] Schram, W. 1954. How communication works. In *The Process and Effects of Communication*. University of Illinois Press, Urbana.

[36] Shannon, C., and Weaver, W. 1949. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana-Champaign.

[37] Sierman, B. 2014. The SCAPE policy framework, maturity levels, and the need for realistic preservation policy. In *Proceedings of the 11th International Conference on Digital Preservation* (Melbourne, October 6-10), 259-266. URL= https://www.nla.gov.au/sites/default/files/ipres2014-proceedings-version_1.pdf.

[38] Stamper, R. 1973. *Information in Business and Administrative Systems*. Wiley, New York.

[39] Svenonius, E. 2000. *The Intellectual Foundations of Information Organization*. MIT Press, Cambridge.

[40] Upward, F. 1996. Structuring the records continuum – Part one: Postcustodial principles and properties. *Archives and Manuscripts* 24, 2 (November), 268-285. URL= http://www.infotech.monash.edu.au/research/groups/rcrg/pub lications/recordscontinuum-fupp1.html.

[41] UC Curation Center. 2015. *Digital Curation Foundations*. URL=http://wiki.ucop.edu/display/Curation/Foundations.

[42] Waters, D., and Garret, J. 1996. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. URL=http://www.clir.org/pubs/reports/pub63.

[43] Wickett, K., Sacchi, S., Durbin, D., and Renear, A. 2012. Identifying content and levels of representation in scientific data. *Proceedings of the ASIS&T 75th Annual Meeting* (Baltimore, October 26-31). URL= https://www.ideals.illinois.edu/handle/2142/35259.

[44] Wimsatt, W., and Beardsley, M. 1946. The intentional fallacy. *Sewanee Review* 54, 468-4.