DRS Futures Project
Conceptual Modeling for Digital Preservation*

Version 1.0 – September 2023

1 Introduction

The Digital Repository Service (DRS) is Harvard University's centrally-supported solution for long-term preservation of its ever-growing digital collections. Harvard Library (HL) began operation of the DRS in October 2000. At that time, no viable commercial or open-source products were available or fully addressed contemporaneous functional requirements. Consequently, it was necessary for HL to build a novel system in-house (Flecker, 2000). Since then, use of the DRS has grown to host over 10.7 million digital objects and 913 million files totaling over 1.8 PB. These materials span all content genres and formats critical to the University's research, teaching, and learning mission as well as its administrative operation. While the DRS technical platform has been continually maintained and incrementally updated over the past two decades (Goethals et al., 2014; Goethals et al. 2015; Goethals & Patterson, 2018; Woods, 2022), it still remains a custom system making increasingly unsupportable demands on finite internal resources. Furthermore, the functional applicability of the DRS is increasingly constrained by limitations arising from long-standing and deep-seated conceptual design, implementation, and operational decisions. In order to address these concerns, HL is now engaged in a needed generational modernization known as the DRS Futures project. This paper describes the philosophical and conceptual foundations of the Futures project, and how they are realized in terms of abstract functional and data reference models.

1.1 DRS Futures Project

The Futures project will revitalize the DRS and reposition it to continue to provide effective, efficient, and sustainable stewardship of the University's digital collections in light of future challenges and opportunities (HL/LTS, 2023). The Futures project is structured in three phases:

1. Envisioning an *ideal* repository
2. Specifying an *achievable* repository
3. Deploying an *operational* repository

The first phase is a purposefully open-ended investigation of aspirational digital preservation needs and goals explicitly unfettered by considerations of how such aspirations ultimately will be provisioned. These ideals will be winnowed down to the achievable in the second phase, contextualized with the aspirational end-goals foremost in mind. As it is rare for any socio-technical innovation to exceed one's ambitions, it is important to set one's aim high and then plan to approach that aim incrementally and

---

* An earlier version of the approach outlined in this paper will be presented at iPRES 2023: Abrams, S., "Rethinking digital preservation: Conceptual foundations," *19th International Conference on Digital Preservation*, University of Illinois, 19-22 September 2023, https://ipres2023.us/timetable/event/tp-2/.

asymptotically over time. In essence, HL is interested in looking beyond what the replacement state-of-the-art might be today, towards what it could – and should – be in the near or far future. Such long-range strategic thinking and planning is possible only when rooted in robust philosophical and conceptual foundations.

This document outlines the philosophical and conceptual principles underlying HL's approach to digital preservation stewardship. Note, however, that these foundational precepts represent aspirational thinking. As such, they are presented for informative rather than prescriptive purposes. In particular, they do *not* constitute formal requirements as expressed in the project Request for Proposal (RFP). However, it is hoped that these ideas *strongly* inform responses to that RFP. While HL is looking for an immediate solution for its digital preservation repository needs, it is also interested in fostering a long-term relationship conducive for rapid, robust, and sustained innovation and advancement in digital preservation infrastructure for its own benefit and that of the wider international digital preservation community. In that regard, current or future consistency or alignment with this conceptual approach outline here will be a positive factor in the evaluation of RFP responses.

1.2 Scope

The DRS is the core component of an evolving ecosystem of technical and service infrastructure (see Figure 1.1) for the effective, efficient, and sustainable long-term stewardship and preservation of the University's deep, broad, rich, and often unique digital collections.

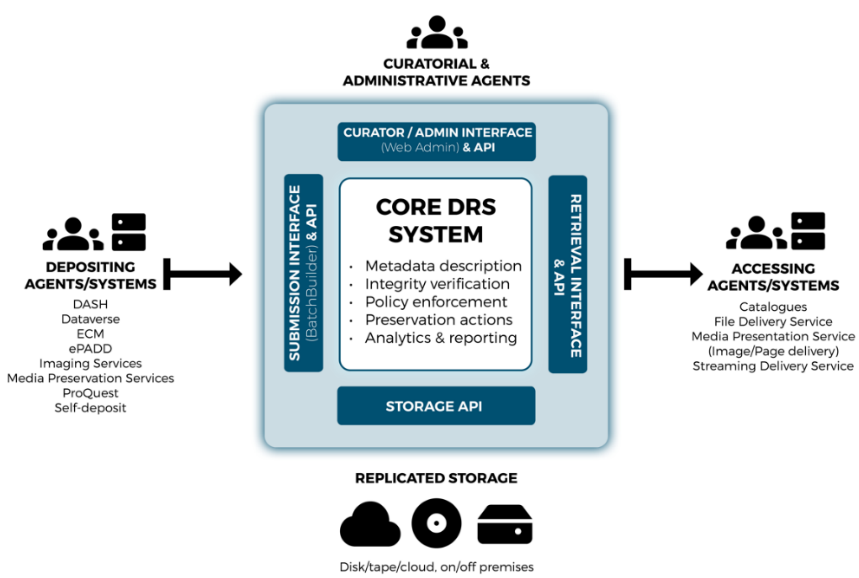

Figure 1.1 – DRS-centered ecosystem

All of these individual components have been carefully designed and implemented to provide a seamless online experience. However, they are subject to distinct product ownership, governance, and support regimes. The Futures project is concerned with modernization of the core DRS system only, although it is imperative that it continues to integrate and interoperate with the surrounding and evolving constellation of added-value systems.

This paper develops a consistent and comprehensive approach to the philosophical and conceptual foundations of the digital preservation enterprise as they can and should be applied to infrastructural design and operation.

2 Exploratory Approach

The process of planning and deploying any significant socio-technical system naturally progresses through distinct stages encompassing initial ideation and subsequent development or procurement (Kneuper, 2017). The transition from the more intangible considerations of the former to the more specific details of the latter is codified in terms of functional and non-functional system requirements. These act variously as a specification for development, an evaluative rubric for procurement, and acceptance criteria for formal project completion. Traditionally, requirements development is approached *inductively* (Jebreen, 2012), relying on extensive stakeholder engagement as well as reference to prior practice, professional intuition, and shared community attitudes to establish needs, goals, and aspirations ultimately refined into a set of use cases (Wiegers & Beaty, 2013).

While both a legitimate and fruitful approach, in order to achieve a higher level of confidence in final requirements, inductive results should be complemented by a parallel *abductive* process that derives requirements from a small axiomatic set of accepted first principles (Danermark et al., 2002; van Lamsweerde, 2008). Abduction is the mode of logical inference that seeks the best possible explanation for a domain's phenomena, in distinction to deduction's logically-necessary and induction's logically-most-probable explanations (Andow, 2016). The final logical refinement of these philosophical and conceptual principles constitutes an abstract reference model (ARM) of the desired system. An ARM is a framework defining the fundamental entities and relationships constituting a domain untethered from the semantics of any specific implementations (MacKenzie et al., 2006).

Due to its logical formality and systematic application, abductive derivation is more likely to result in comprehensive coverage of appropriate domain considerations relative to a more ad hoc and anecdotal inductive process, however well-grounded it may be in historical precedent, domain best practice, and professional experience. In essence, the top-down abductive approach starts with a high-level model of the entire domain under consideration and systematically segments it into smaller and smaller units of greater and greater conceptual detail. The bottom-up inductive approach, on the other hand, starts with various granular units of detail that are gradually refined and abstracted with an assumption that they will eventually cohere into comprehensive coverage of the full domain. Ideally, the two approaches will exhibit significant, if not full, overlap. Regardless, performing the two activities in parallel provides an opportunity to identify and fill in any gaps resulting from the individual exercises.

3 Philosophical Inquiry

The foundational basis for the Futures project emerged through a process of Philosophical Inquiry (PI). PI is a qualitative research method that derives meaning from experience through abductive questioning of fundamental assumptions within a domain of practice to propose new, and better, explanatory structures for that domain (Burbules & Warnick, 2006; Grace & Perry, 2013). In the Futures context, the abductive inquiry began with questioning the fundamental nature of the digital preservation enterprise. The *Encyclopedia of Archival Science* defines digital preservation as “the processes and controls that enable digital objects to survive over time” (Thibodeau, 2017). This formulation – broadly representative of historic and current domain thinking – emphasizes an object- and process-centric view that implicitly promotes a metaphoric narrative of digital preservation as a *managerial* endeavor. That is, a set of activities done *to* objects to ensure persistence of their significant characteristics over time. While an important foundational step, this narrative minimizes critical attention to what subsequently can be done *with* those objects and to what *effect*.

Similarly, the phraseology common to other community-accepted definitions of the preservation field – for example (all emphasis added): “policies, strategies, and *actions* that ensure *access* to digital content over time”(ALA, 2009); “act of *maintaining* information, independently Understandable by a Designated Community, and with evidence supporting its Authenticity, over the Long Term” (ISO, 2012); “series of *managed* activities necessary to ensure continued *access* to digital materials for as long as necessary” (Beagrie, 2015); “*processes* aimed at ensuring the continued *accessibility* of digital materials” (UNESCO, 2019) – emphasizes two points. First, that the primary agential role within the domain is an enabling instrumental one, e.g., acting as strategiz*ers*, maintain*ers*, manag*ers*, process*ers*, Second, that the imperative goal of the exercise is provision of artifactual access. Access refers to the ability and permission to find and retrieve information relevant for a specific purpose (SAA, 2020). In other words, access is an *enabling* factor for subsequent use, which remains a distinct phenomenon. While this distinction may be operationally prudent (Wilson, 2017), it is teleologically problematic. The consensual weight of repeated assertions of the *operational* primacy of accessibility implicitly positions digital preservation *conceptually* as an essentially managerial activity, whose imperatives stop with provisioning access (Abrams, 2021). The ability to retrieve a well-managed object, however, is distinct from a subsequent ability to make productive use of it. The parameters of that usage are concerned with post-managerial experience.

3.1 Digital Preservation as Communicative Endeavor

The embrace of that experiential component recasts digital preservation as an essentially *communicative*, rather than merely managerial enterprise. That is, it aims to facilitate future purposive human engagement with past informative expression. While that facilitation necessarily involves technological intermediation through artifactual vehicles and managerial processes, its teleological imperatives are fundamentally humanistic in nature. These give preeminence to the role of the ultimate information consumer (Rogala & Bialowas, 2017; Sacchi, 2017) and the communicative outcomes of the consumer/content engagement. The success of such an act of preservation-enabled communication is dependent on its cognitive, affective, and conative consequence. That is, preservation acts are successful if they result in something pertinent to the use at hand being newly known intellectually, felt psychologically, or performed physically that would not otherwise have occurred (Kuhlthau, 2017; Savolainen, 2019). As any such use is contingent with respect to time, place, person, and purpose

(Bishop & Hank, 2018; Morrissey, 2014), preservation-enabled communication inherently operates in an intersubjective sphere. Efforts to ensure beneficial outcomes over time are complicated by the fact that the passage of time is inexorably accompanied by ever-growing technical distance. However, the more significant preservation challenge over archival timespans is the accompanying *cultural* distance separating the points of content creation, acquisition, and use.

It is important to note that while preservation is an act of *communication*, it is not fundamentally a *conversation*. That is, it represents a unidirectional, rather than bi-lateral, transfer of information from the past into the future. Unlike a conversation, there is no practical means of asking the past for clarification or restatement. This inherent aspect of the preservation enterprise emphasizes the importance of capturing and persisting the richest possible information set and ancillary interpretive context.

Concern	Managerial				Communicative		
Referent	Information object				Information experience		
Focus	Artifactual				Experiential		
Abstraction	Carrier	Message			Performance	Environment	Mind
Function	Reifactory	Representational	Rhetorical	Ontological	Epistemological	Associational	Phenomenological
Affordance	Manifestation	Encoding	Expression	Meaning	Behavior	Context	Understanding
Semiotic	Ontics	Empirics	Syntactics	Semantics	Performics	Plaistics	Pragmatics
Imperative	Integrity	Validity	Authenticity	Reliability	Accessibility	Relevancy	Legitimacy
Descriptiveness	Is-ness		Of-ness			About-ness	
Role	Instrumental (enabl*ing* means)				Teleological (enabl*ed* ends)		
Measure	Quantitative output				Qualitative outcome		
Metric	Trustworthiness				Success		
Evaluation	Objective				Intersubjective		

Table 3.1 – Philosophical Foundations of Digital Preservation

3.2 Digital Preservation as Semiotic Activity

A communicative perspective on the digital preservation domain makes it susceptible to a communicological approach to responsive policy, strategy, and tactics. Communicology is the study of individually-embodied human discourse (Lanigan, 2015), in distinction to disembodied machine-to-machine information-theoretic communication (Lanigan, 2013) and socially-embodied mass communication (Catt, 2014). That discourse is viewed as a semiotic system of expressive signs whose meaning emerges through contingent interpretation by a consumer individually, institutionally, and culturally-positioned in socio-technical space and time (Nöth, 1990; Mingers, 2017). A "sign" is high-level communicative abstraction for any information-laden entity that "stands to somebody for something in some respect or capacity" (Peirce, 1932). Stamper extended the traditional tripartite consideration of a sign – syntactic form, semantic meaning, pragmatic experience (Morris, 1964) – to encompass six aspects pertinent for greater applicability to digital information systems (Stamper, 1993) (see Table 3.1). (Stamper's "social" aspect is renamed "plaistic" [contextual] for consistency of notation.) A seventh, performic, aspect is necessary for pertinence to digital preservation (Abrams, 2023). This recognizes that digital objects must be dynamically and contextually *performed* to be susceptible to analog human perception and cognitive interpretation (Heslop et al., 2002; Becker, 2018).

The common metaphor of a digital *carrier* is the ontic (or tangibly-reified) manifestation of an abstract information-laden *message*. That message encompasses three distinct semiotic aspects:

1. Empiric symbolic encoding
2. Syntactic rhetorical expression
3. Semantic cognitive or affective meaning

These generally align with the FRBR Manifestation, Expression, and Work constructs (Riva et al., 2016), which constitute an essential progression from the (relatively) concrete to the (relatively) abstract.

The seven semiotic dimensions correspond to a set of functional categories providing associated affordances:

1. *Reifactory* – Concrete manifestation of abstract informational content
2. *Representational* – Digital encoding underlying that reifaction
3. *Rhetorical* – Persuasive intellectual/aesthetic structure of that encoding
4. *Ontological* – Meaning or affect of that rhetorical structure
5. *Epistemological* – Modality of engaging with that ontological meaning
6. *Associational* – Individual/institutional/social context in which that modality is exercised

The semiotic dimensions similarly correspond to a continuum of perspectives on the preservation enterprise from the objective to the intersubjective, spanning three descriptive categories:

1. Constitutive *is-ness*
2. Denotative *of-ness*
3. Connotative *about-ness*

This terminology is borrowed from subject cataloging theory (Hjørland, 2017), but is deployed here to indicate the range of afforded semiotic descriptive scope. For example, while this paper *is* an Office Open XML document, it is overtly descriptive *of* the derivation and application of a conceptual domain model aiding the Harvard Library's infrastructure refresh, while also being interpretatively *about* the novelty and legitimacy of that model as a complement to prior modeling efforts.

3.3 Preservation Success Through a Multi-Valent Lens

Preservation outputs and outcomes can be evaluated in terms of associated imperative qualities. An output is a quantifiably-measurable result of an activity, such as counts or enumerations of the generated states or productions of a system or process (Dugan & Hernon, 2022), while an outcome is a qualitatively-assessable benefit of an output (Kyrillidou, 2002). Thus, an outcome focuses on the experiential impact or difference an output has on the part of its recipient (Tsakonas & Papatheodorou, 2011). An ontic manifestation is *integral* if it is complete and uncorrupted (Kastenhofer, 2015); an empiric encoding is *valid* if it conforms to an authoritative definition (Lindlar & Tunnat, 2017); a syntactic expression is *authentic* if it expresses what it purports to express (Kastenhofer, 2015); a semantic meaning is *reliable* if its factual presentation is accurate (Kastenhofer, 2015); a performic behavior is *accessible* if it can be availed upon at a time and place and in a manner of the consumer's choice (Jaillant, 2022); a plaistic context is *relevant* if it is fit for the consumer's intentional or serendipitous purpose (Lee, 2011); and a pragmatic understanding is *legitimate* if it is meaningful for that purpose (Dallas, 2007; Duranti, 2006). Since any given encounter with preserved digital material is contingent with respect to time, place, person, and purpose, the consuming participant in that encounter will come to it with a potentially unique set of implicit or explicit weighting factors regarding the relative importance of these various qualities. Thus, digital preservation success should be viewed as a multi-valent evaluable factor.

4 Conceptual Foundations

The success of long-term digital preservation activity is commonly evaluated in terms of four normative qualities: the integrity, authenticity, accessibility, and usability of managed digital resources. Authenticity can be seen as subsuming integrity. That is, any explicit loss of integrity inherently implies corresponding loss of authenticity. These evaluative factors emerge through content analysis of digital preservation policy documents objects (Abrams, 2021). Given that such policies establish the parameters of the implicit, but nevertheless controlling, social "contract" underlying the interaction between preservation stakeholders and service-providers, whether internal or external to an institutional program, these qualities suggest three defining imperatives for the preservation enterprise:

1. Ensuring persistence of *authentic* information *objects*
2. Supporting modalities of *authoritative* information *performances*
3. Affording opportunities for *legitimate* information *experiences*

In this regard, authenticity is the quality of an object being what it purports to be; authoritativeness, the quality of being appropriate and reliable for the purpose at hand; and legitimacy, the quality of being meaningful for that contextually-situated purpose. (Authenticity is viewed as subsuming integrity, as any explicit loss of integrity inherently implies corresponding loss of authenticity.) These correspond to three perspectives on preservation persistence along ontological, epistemological, and phenomenological dimensions, respectively. That is to say, there is an intention and expectation that future preservation outcomes encompass the preserved artifact *itself*, the preserved means to interact with and *know about* the artifact, and the preserved experiential *impact* of that interaction.

The authenticity/legitimacy dichotomy is between objective universality (authenticity) and intersubjective contingency (legitimacy). In other words, while a given digital object is singularly either authentic or inauthentic, that same object may be susceptible to any number of legitimate (re)uses, each particular to some socio-technically-situated time, place, person, and purpose. Efforts to ensure these beneficial outcomes over time is complicated by the ever-increasing number, size, complexity, and diversity of digital content available for long-term stewardship, as well as the continual – and often disruptive – evolution and transformation of the modalities of desired (re)use.

4.1 Affordance and Context in Digital Preservation

The perspectival shift in digital preservation emphasis towards communicative information experiences suggests the desirability of similarly recasting the domain concept of significant properties to that of significant *affordances* (Abrams, 2023). In the preservation context, an affordance is a functional capability available to a human consumer *to* do something meaningful *with* a preserved object (Ossher & Tarr, 2022). For example, the property of fixity *affords* the ability to determine integrity. Similarly, the property of an image's defined color space *affords* the ability for colorimetrically-reliable visual presentation. In other words, an affordantial perspective complements a focus on the managerial and artifactual aspects of preservation attention with communicative and experiential considerations. The experiential connotation of affordance also highlights the view of human engagement with a preserved digital object as an intersubjective performance (Becker, 2018). The meaningfulness of the pragmatic response to such a performance is dependent on various frames-of-reference that contextualize the encounter (Dappert & Farquhar, 2009). These include the contexts of (Abrams, 2015):

1. Cultural production, indicative of originating creative intention
2. Curatorial appraisal, selection, and aggregation in thematic collections, through which the individual member objects accumulate associational meaning (Bonn et al., 2016)
3. Prior consumption, indicative of alternative interpretive reception and response
4. Collateral lived-experience and proximate purpose of the contemporary consumer, establishing teleological expectation

While the domain concept of representation information is defined in generic terms (ISO, 2012), in practice it has not encompassed the means to represent, capture, and retain all of these diverse contextual positions (Brocks et al., 2010). New infrastructural systems should provide explicit support for persistent management of and experiential access to authoritative performative behaviors and relevant contextual reference frames.

4.2 Emergent Programmatic Imperatives

Digital preservation is a complex of people, policies, procedures, as well as systems facilitating technologically-mediated but fundamentally human communication across time (Abrams, 2019). Given that technical infrastructure is inherently ephemeral and needs to be refreshed and re-envisioned periodically (Janée et al., 2009; Barateiro et al., 2010), it is appropriate to assert expansive aspirations for its function and operation during its ideational phase. While these may not be immediately provisionable, they set a benchmark for incrementally-achievable programmatic goals. For the Futures project, these goals include effective, efficient, and sustainable support for:

1. Any content ...
    a. *Genre*, *language*, *form*, or *structure*
    b. *Number* or *size*

   c. *Description* (or none)

2. Any managerial ...
   a. *Duration* (interim, persistent, permanent)
   b. *Eventuality* (proactive whenever possible, reactive whenever necessary)
3. Any stakeholder ...
   a. *Competency*
   b. Purpose
   c. Modality

The first aspirational grouping is concerned with maximizing the scope of preservation eligibility; the second, the range of preservation intentions and expectations; and the last, the parameters of experiential (re)use. A claim of effective support for these various goals does not necessarily imply a uniform level of outcome. Instead, effectiveness should be viewed as the condition of doing the best one can regarding a given body of digital content at a particular point in time and contemporaneous state of expertise, tooling, and capacity as well as controlling curatorial priority.

5 Abstract Reference Modeling

An abstract reference model (ARM) is a useful conceptual benchmark for planning the procurement or development of a socio-technical system at both functional and data-representational levels.

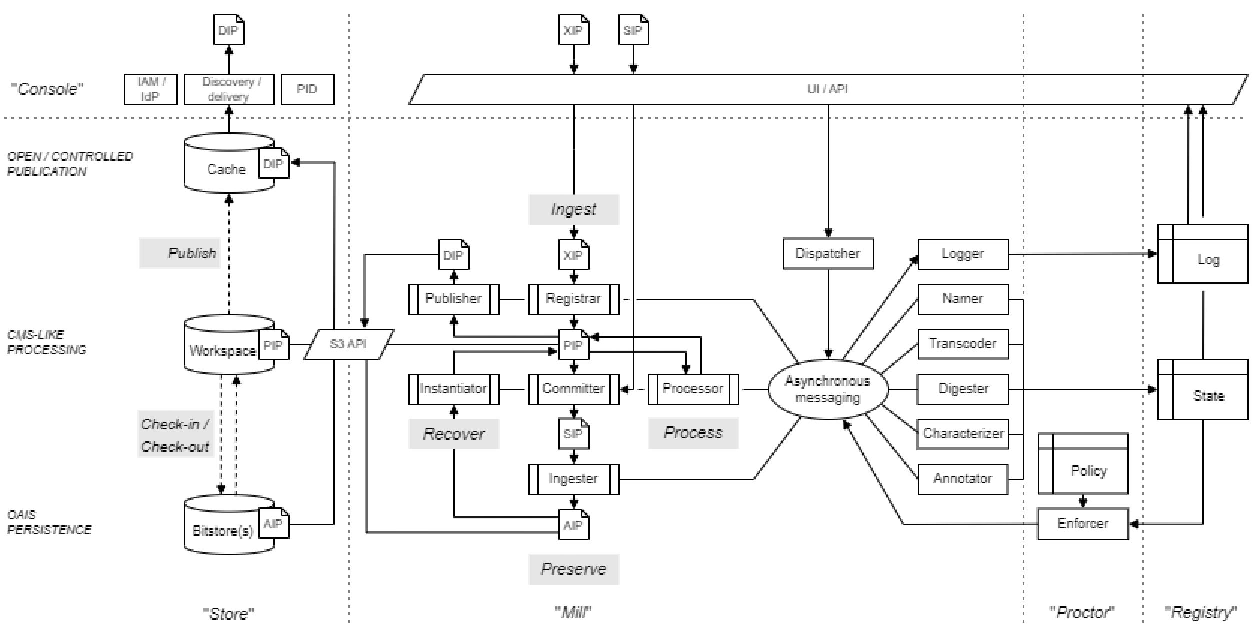

Figure 5.1 – Functional reference model

5.1 Functional Reference Model

The Futures functional model encompasses computational structures for storage, processing, state, and agential interaction, all choregraphed through asynchronous messaging and controlled through automated enforcement of policy-compliant state transitions (see Figure 5.1). These structures conceptually coalesce into five high-level component set abstractions spanning concerns for artifactual and communicative imperatives at informational and procedural levels (see Table 5.1):

1. *Console* – Interfaces for human and automated agential interaction
2. *Registry* – Persistence of empiric, syntactic, semantic, performic, and plaistic state for hosted digital content and logging of infrastructural processes
3. *Proctor* – Machine-actionable policies and automated enforcement
4. *Mill* – A microservice-based processing farm ensuring policy-consistent state
5. *Store* – Ontic, or bit-level, persistence of the tangible manifestations of hosted content

(The lower two sets are named in playful allusion to the Babbage/Lovelace Analytic Engine (Dickey, 1987).) Note, again, that this is an *abstract* description of core functional entities and relationships. Pointedly, it is *not* intended directly as an architectural diagram or technical specification.

<table>
<tr><th></th><th colspan="2">Artifactual</th><th></th><th colspan="2">Communicative</th></tr>
<tr><td>Informational</td><td>Ontic state</td><td>“Store”</td><td rowspan="2">Imperative enforcement
“Proctor”</td><td>Higher-level semiotic state</td><td>“Registry”</td></tr>
<tr><td>Procedural</td><td>Persistence-assuring processes</td><td>“Mill”</td><td>Interactive modalities</td><td>“Console”</td></tr>
</table>

Table 5.1 – Abstract component sets

Diagrammatically, the Registry, Proctor, Mill, and Store sets are presented along the horizontal axis in Figure 5.1. For visual convenience, the Console set spans the horizontal axis.

The model also defines three functional tiers, presented along the vertical axis:

1. *Open/controlled publication* – Discovery and delivery subject to appropriate access control
2. *CMS-like processing* – For initial or subsequent curatorial (re)processing
3. *OAIS-like persistence* – For assured long-term preservation

The four imperative digital preservation norms – integrity, authenticity, accessibility, usability – present a stepwise hierarchy of increasing added-value assurance and utility. The foundational normative basis is provision for bit-level integrity of preserved materials. Without provision for reliable hosting and retrieval of the tangible manifestations of digital objects, none of the higher-level functional imperatives can be addressed competently and completely.

5.1.1 The Store

The Store provides assurance of primary ontic state through the dynamic replication of digital content across a range of technically, geographically, and modally heterogeneous storage platforms. That is, an explicit reliance of multifarious media, platforms, and vendors; locations and threat profiles; and operating conditions, e.g., online/nearline/offline. Disposition of a given unit of digital content to some subset of these platforms is dependent on alignment of curatorial expectations with media and modal platform characteristics. This policy-driven disposition is enabled by standardization on a common access protocol/API and media structuring layout (Woods, 2022) as well as availability of a mechanism for dynamic policy enforcement.

There are three primary components of the Store:

1. *Cache* – High-performance storage optimized for access delivery
2. *Workspace* – High-transactionality storage optimized for thread-safe processing
3. *Bitstore*(*s*) – High-density storage optimized for (globally) reliable persistence

5.1.2 The Mill

The core of the Futures functional model conceives of ideal digital preservation infrastructure as a finite state machine (FSM) (Wagner et al., 2006). Stateful transitions are initiated by either external or internal stimuli, that is, user-specified requests such as new deposit submissions, or self-identified conditions such as fixity violations.

The Mill is composed of sets of workflow swimlanes (Gadatsch, 2023) and microservices (Abrams et al., 2011). Workflows are defined by an extended set of relevant information packages (ISO, 2012):

1. *XIP* – External information package (pronounced “zip”): Arbitrary content structure for external contributed form
2. *PIP* – Processing information package: Content structure for interim processable form in the Processing tier
3. *SIP* – Submission information package: Content structure for final processed form
4. *AIP* – Archival information package: Content structure for internal preservable form in the Persistence tier
5. *DIP* – Dissemination information package: Content structure applicable for external open or controlled access in the Access tier

The AIP, DIP, and SIP package types were introduced as part of the ISO 14721 OAIS information model. The PIP and XIP types are newly defined for convenient and consistent reference within the ARM. The workflow swimlanes are defined in terms of the significant package transformations underlying preservation stewardship:

1. *Registrar* – Converting XIPs to PIPs, for unprocessed materials accessioned into the system
2. *Committer* – Converting PIPs to SIPs, for checking-in processed material from the Processing layer into the Persistence layer
3. *Ingester* – Converting SIPs to AIPs
4. *Processor* – Converting PIPs to PIPs, reflecting curatorial processing
5. *Instantiator* – Converting AIPs to PIPs, for checking-out material from the Persistence layer to the Processing layer for subsequent curatorial (re)processing or disaster recovery

6. *Publisher* – Converting PIPs to DIPs, for publishing processed material for discovery and delivery

5.1.3 The Proctor

The Proctor performs automated policy enforcement by evaluating stimuli in light of current content state and applicable policy rules (Görzig et al., 2016; Conway, 2017). If necessary, the Proctor dispatches a series of potentially chained microservice invocation requests intended to bring the state back into conformance with policy prescriptions. IRODS provides a useful exemplar in this regard (Conway et al., 2011; Conway, 2017). The Preservation Action Registries (PAR) initiative (O'Sullivan & Tilbury, 2021) suggests an avenue of exploration regarding the expression and evaluation of policy rules.

5.1.4 The Registry

The Registry encompasses persistent structures of primary communicative state, i.e., at the empiric, syntactic, semantic, performic, and plaistic levels. Principles and requirements for representing this state is introduced as part of the information reference model in Section § 5.2. For purposes of disaster recovery, all stateful information is persisted in the Store tier, which is considered the copy-of-record. The Registry representation should be viewed as an optimization for performant indexing and searching of state.

5.1.5 The Console

Interaction with the repository takes place through functionally-equivalent command line interfaces (CLI), online user interfaces (UI), and programmatic APIs.

5.2 Information Reference Model

Information modeling lies at the heart of the ARM. While preservation infrastructure and functional processes are inherently ephemeral and ultimately expendable, the persistence of digital content itself is the core programmatic imperative (Abrams et al., 2009). The core design principle for the information reference model is to be fully self-describing and self-preserving. That is, all information elements of both contributed content and the system itself are treated as primary preservable resources, making the entire repository reinstantiatable from its preserved representation.

The DRS reference data model (see Figure 5.2) draws upon several significant antecedent modeling efforts, incorporating their successful elements, avoiding their pitfalls, and addressing their lacunae:

1. The Digital Object Architecture (DOA) (Kahn & Wilensy, 2006) assumes a fundamental distinction between modeled *data* and *key-metadata*.
2. The Open Archival Information System (OAIS) Reference Model (ISO, 2012) extends the concept of object metadata into a recursive *representation information* structure.
3. The Metadata Encoding & Transmission Standard (METS) (Library of Congress, 2022) distinguishes between the physical and logical structures of digital content.

4. The Portland Common Data Model (PCDM) (DuraSpace, 2018) defines a shallow hierarchy of collections, objects, and files. PCDM's role as a common interoperability and exchange format necessitates a very high-level and generic design that does not easily distinguish between conceptual and concrete aspects of modeled entities or provide fine enough granularity of representational detail to support long-term preservation analysis, planning, and intervention.
5. The Hydra::Works data model (Samvera, 2018) extends PCDM to include the concept of a FileSet as an intermediate layer between Object and File. However, the model provides little guidance on the criteria for or appropriate relationships between members of a FileSet or distinct FileSets.
6. The E-Ark common information package specification (Bredenberg et al., 2016) makes important distinctions between representational, descriptive, ancillary, and documentary expressions.
7. Jisc Open Data Hub canonical data model is optimized for the representation of research data, although that genre is considered expansively (Jisc, 2018; Stokes et al., 2019).
8. The PREMIS data dictionary for preservation metadata was purposely developed to support preservation imperatives and is broadly supported by infrastructural products (Library of Congress, 2015). It also introduces the Hydra FileSet-like concept of a Representation.
9. The IFLA $FRBR_{OO}$ model (Bekiari et al., 2015), the object-oriented derivative of the original entity-relationship-based $FRBR_{ER}$ model (Riva et al., 2017), explicitly addresses the important distinction between abstract and concrete considerations of modeled content.
10. The CIDOC conceptual reference model (CRM) (ICOM, 2022) is a comprehensive ontology for the cultural heritage and museum domains.

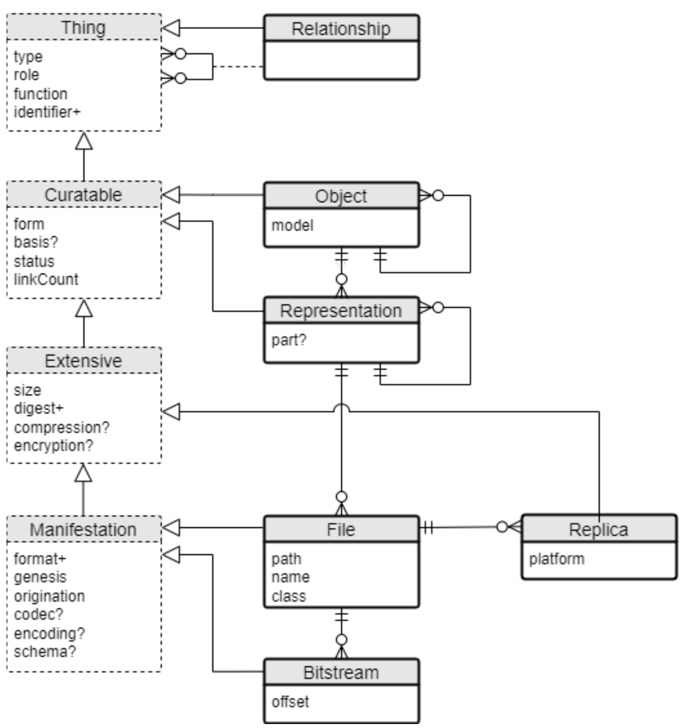

Figure 5.2 – Information modeling hierarchy

5.2.1 Thing

A Thing is the non-instantiable conceptual parent from which all other modeled entities implicitly inherit. A Thing is anything with persistent independent identity distinguishing it from all other possible things, and thus, cognate with a FRBR Res (Riva et al., 2016) (see Tables § 5.2 and 5.3).

ENTITY Thing			
type	1..1	ENUM	Structural type: What the Thing *is*
role	1..1	ENUM	Purposive role: What the Thing *does*
function	1..1	ENUM	Informative function: What value the Thing *affords*
identifier	1..*	Identifier	Persistent identifier

Table 5.2 – Thing entity

PROPERTY Identifier			
namespace	1..1	ENUM	Namespace
value	1..1	STRING	Value

Table 5.3– Digest property

Type, role, and function are a uniform set of descriptive characteristics applicable to all modelled Things. The enumerated ranges of permissible types, roles, and functions for various Things are context sensitive. These characteristic properties are important for purposes of entity distinction, classification, aggregation, and self-description.

5.2.2 Relationship

A Relationship is an instantiable Thing defining a typed association between two arbitrary Things (see Table 5.4; dashed cell outlines indicate inherited entity properties). Relationship functions are conventionally labelled from the perspective of the target rather than source Thing, e.g., “function: IsPartOf” rather than “function:HasPart”.

ENTITY Object IS-A Thing			
...	...	...	...

Table 5.4 – Relationship entity

5.2.3 Curatable

A Curatable is non-instantiable Thing that is the direct object of preservation attention (see Table 5.5).

ENTITY Curatable IS-A Thing			
form	1..1	ENUM	Expressive form: How the Curatable is *organized*
basis	0..1	ENUM	Morphological basis: *Structuring* principle of that organization
status	1..1	ENUM	Status: Active — Active (default) Deleted — Logically deleted, but recoverable Purged — Physically deleted
linkCount	1..1	INTEGER	Reference link count; default = 0

Table 5.5 – Curatable entity

The enumerated ranges of permissible forms and bases for various Curatables are context sensitive. These characteristic properties are important for purposes of entity distinction, classification, aggregation, and self-description.

5.2.4 Object

An Object is an instantiable Curatable that is the fundamental unit of repository *intellectual* management, encapsulating some coherent unit of informative content, and thus, cognate with a FRBR Work (Riva et al., 2016) or PREMIS Intellectual Entity (Library of Congress, 2015) (see Table 5.6). As opposed to a Representation, an Object is assumed to have an independent curatorially-meaningful intellectual identity.

ENTITY Object IS-A Curatable			
type	1..1	ENUM	Structural type: Multipart — Hierarchically-nested Object Simple — Stand-alone Object
role	1..1	ENUM	Purposive role: Content — Object of curation interest System — Object of operational convenience
…	…	…	…
model	1..1	ENUM	Object content model

Table 5.6 – Object entity

5.2.5 Representation

A Representation is a non-instantiable Curatable that realizes of some constituent component of a Object's information content, and thus, cognate with a FRBR Expression or PREMIS Representation (see Table 5.7). As opposed to an Object, a Representation is assumed to a technically significant informative component, but without an independent curatorially-meaningful intellectual identity.

ENTITY Representation IS-A Curatable			
type	1..1	ENUM	Structural type: Granular Hierarchically-nested Representation Singular Stand-alone Representation
Role	1..1	ENUM	Purposive role: Substantive Primary content Descriptive Descriptive content Instrumental Instrumental content
…	…	…	…
part	0..1	ENUM	Representational part

Table 5.7 – Representation entity

5.2.6 Extensive

An Extensive is a non-instantiable Curatable with properties dependent on its specific concretization extensive in (digital) space (see Table 5.8).

ENTITY Extensive IS-A Curatable			
…	…	…	…
size	1..1	INTEGER	Size in octets (8-bit bytes)
digest	0..*	Digest	Message digest (or hash or checksum)
compression	0..1	ENUM	Compression algorithm default: none
encryption	0..1	ENUM	Encryption algorithm; default: none

Table 5.8 – Extensive entity

5.2.7 Manifestation

A Manifestation is a non-instantiable Extensive that materializes some constituent component of a Representation's information content, and thus, cognate with a FRBR Manifestation (see Tables § 5.9 and 5.10).

ENTITY Manifestation IS-A Extensive			
…	…	…	…
format	0..*	ENUM	Format
codec	0..1	ENUM	Optional codec for audio/video manifestations
encoding	0..1	ENUM	Optional character encoding for textual manifestations; default: UTF-8
schema	0..1	ENUM	Optional metadata schema for metadata manifestations
genesis	1..1	ENUM	File genesis: Born-digital Digitized (default) Unknown
origination	1..1	ENUM	File origination:

			Corrected Derived Normalized Original (default) Processed Other Unknown

Table 5.9– Manifestation entity

PROPERTY Digest			
algorithm	1..1	ENUM	Message digest algorithm
value	1..1	STRING	Lower-case hexadecimal message digest value

Table 5.10– Digest property

Manifestation forms are specified in terms of MIME top-level media types (Freed et al., 2013; IANA, 2023).

5.2.8 File

A File is an instantiable Manifestation that is the fundamental unit of repository *data* management, consisting of a formatted logical byte sequence with file system semantics (IEEE, 2018), and thus, cognate with a PREMIS File (see Table 5.12). A File is instantiated by an arbitrary number of Replicas and may consist of an arbitrary number of internal Bitstreams.

ENTITY File IS-A Manifestation			
type	1..1 1..1	ENUM	Structural type: Container Arbitrary aggregation Wrapper Relational aggregation Unitary Stand-alone File
role		ENUM	Purposive role: Data Primary data Metadata Descriptive metadata
...	...	...	...
path	1..1	STRING	Original pathname
name	1..1	STRING	Original filename
class	1..1	ENUM	File storage classification: AR Archival BA Basic DA Data DE Delivery EX External LG Large RE Repurposable SE Sensitive

Table 5.11 – File entity

5.2.9 Replica

A Replica is an instantiable Extensive specific retrievable *copy* of a File as manifest on some persistent storage medium, and thus, cognate with a FRBR Item (see Table 5.13).

ENTITY Replica IS-A Extensive			
...	...	...	...
platform	1..1	ENUM	Storage platform

Table 5.12 – Replica entity

5.2.10 Bitstream

Arbitrary subset of a File's logical byte sequence, and thus, cognate with a PREMIS Bitstream (see Table 5.14).

ENTITY Bitstream IS-A Manifestation			
...	...	...	...
offset	1..1	INTEGER	Octet (8-bit byte) offset from octet 0 of containing File

Table 5.13 – Bitstream entity

5.3 System Objects

A system object encapsulates information pertinent to the operation of the repository itself. The objects encapsulating this information are subject to normal preservation oversight and outcome, including persistence. Having them available in the storage persistence layer (the "store") ensures the sustainability of the system in the aftermath of catastrophic system failure, as the full state of the system can be reinstantiated from the store.

1. *Actor* – Human stakeholders with user privileges regarding the repository
2. *Agent* – Automated processing components of the repository
3. *Charset* – Character encoding scheme
4. *Codec* – Empiric rules for internal audio/video File organization
5. *Collection* – Aggregations of Objects on a topical or thematic basis
6. *Compression* – Algorithmic method for removing redundancy from data representations
7. *Digest* – Algorithmic one-way fixed length message digest or "hash"
8. *Encryption* – Algorithmic method for transforming data into a secure ciphertext
9. *Format* – Ontic, empiric, and syntactic rules for internal File organization
10. *Identifier* – Identifier namespace
11. *Model* – Ontic, empiric, syntactic, and semantic rules for internal Object organization and external relationships

12. *Objective* – Specific achievable result of preservation attention (CCSDS, 2019)
13. *Schema* – Syntactic and semantic rules for expressive characterizing metadata
14. *Storage* – Hardware/software environment for data persistence

6 Conclusion

The foundational conceptualization of a domain establishes the metaphoric as well as pragmatic boundaries of legitimate domain focus and action (Condon, 2014). Current perspectives of the digital preservation enterprise promote a view largely limiting its concerns to the managerial and artifactual. While these are necessary enabling factors, they do not encompass sufficient attention to the communicative and experiential aspects of preservation concern. Fuller understanding and exploitation of the domain follows from complementary attention to both the instrumental means as well as the teleological ends of the enterprise. The latter can be summarized as facilitating system-mediated but fundamentally human communication unfolding across archival timespans and accompanying technical and cultural distance. Progress towards this goal revolves around three primary digital preservation imperatives: ensuring persistence of authentic information objects; providing authoritative information access modalities; and affording opportunities for legitimate information experiences. Considerations pertinent to the first, ontological imperative are well-examined and modeled at the abstract (ISO, 2012), architectural (Tallman, 2021), and deployment (DPC, 2022) levels. Similar efforts regarding the second, epistemological imperative are emerging through research and practice in software preservation and emulation (Cochrane et al., 2022). Intentions and practices supporting the third, phenomenological imperative are less mature. The communicological framework proposed here provides useful structuring principles for further investigation of this final – and teleologically preeminent – concern.

The Harvard Library DRS Futures project team used this communicological framework as the basis for an open-ended abductive exploration of the constitutive components of an ideal digital preservation infrastructure. This process derived a novel abstract reference model for that infrastructure from a set of initial axiomatic principles. While the contours of the model infrastructure are unlikely to be fully provisionable in the near term, they nevertheless constitute a critical roadmap for long-term planning of the Library's digital preservation intentions. A subsequent phase of the Futures project will derive a constrained version of the idealized vision that is achievable and ultimately procurable and deployable. In almost all socio-technical endeavors, it is very unlikely that achievement ever exceeds aspiration. Consequently, it is reasonable to set high aspirations as a benchmark for a desirable goal that can be approached asymptotically. The Library hopes that its efforts to provide a new conceptual foundation for digital preservation thinking will contribute not only to the success of its institutional stewardship priorities, but also to progress in community-wide state-of-the-art innovation and adoption.

References

Abrams, S., Cruse, P., & Kunze, J. (2009). "Persistent objects, disposable systems," *4th International Conference on Open Repositories*, Atlanta. http://hdl.handle.net/1853/28490

Abrams, S., Cruse, P., Kunze, J., & Minor, D. (2011). "Curation micro-services: A pipeline metaphor for repositories," *Journal of Digital Information* 12(2). https://journals.tdl.org/jodi/index.php/jodi/article/view/1605/1766

Abrams, S. (2015). "A foundational framework for digital curation: The Sept domain model," *12th International Conference on Digital Preservation*, Chapel Hill. https://hdl.handle.net/11353/10.429533

Abrams, S. (2019). "Theorizing success: A communicological approach to evaluating digital preservation efficacy," *Bulletin of IEEE Technical Committee on Digital Libraries* 15(1). https://bulletin.jcdl.org/Bulletin/v15n1/papers/abrams.pdf

Abrams, S. (2007). "File formats," *Digital Curation Reference Manual*, eds. S. Ross & M. Day (Edinburgh: Digital Curation Center). https://www.dcc.ac.uk/resources/curation-reference-manual/completed-chapters/file-formats

Abrams, S. (2021). "Tacit attitudinal principles for evaluating digital preservation success," *Archival Science* 21(3): 295-315. doi:10.1007/s10502-021-09360-5

Abrams, S. (2023). *A Communicological Critique of Evaluative Norms for Digital Preservation Success*, PhD dissertation, Queensland University of Technology. doi:10.5204/thesis.eprints.238194

ALA [American Library Association] (2009). *Definitions of Digital Preservation*. http://www.ala .org/alcts/resources/preserv/defdigpres0408

Andow, J. (2016). "Abduction by philosophers: Reorienting philosophical methodology," *Metaphilosophy* 47(3): 353-370. doi:10.1111/meta.12191

Barateiro, J., Antunes, G., Freitas, F., & Borbinha, J. (2010). "Designing digital preservation solutions: A risk management-based approach," *International Journal of Digital Curation* 5(1). doi:10.2218/ijdc.v5i1.140

Beagrie, N., ed. (2015). *Digital Preservation Handbook*, 2nd ed., Digital Preservation Coalition. https://www.dpconline.org/handbook

Becker, C. (2018). "Metaphors we work by: Reframing digital objects, significant properties, and the design of digital preservation systems," *Archivaria* 85. https://archivaria.ca/index.php/archivaria/article/view/13628

Bekiari, C., Doerr, M., Le Bœuf, P., and Riva, P. (2015). *Definition of FRBROO: A Conceptual Model for Bibliographic Information in Object-Oriented Formalism*, Version 2.4, IFLA. https://www.ifla.org/publications/node/11240

Bishop, B. W., & Hank, C. (2018). "Measuring FAIR principles to inform fitness for use," International Journal of Digital Curation 13(1): 35-46. doi:10.2218/ijdc.v13i1.630

Bonn, M., Kendall, L., and McDonough, J. (2016). "Preserving intangible heritage: Defining a research agenda," *Proceedings of the Association for Information Science and Technology* 53(1): 1-5. doi:10.1002/pra2.2016.14505301009

Bredenberg, K., Skog, B., Nielsen, A. B., Johansen, K. H. E., Tømmerholt, P., Thirifays, A., Schlarb, S., Rörden, J., Wilson, A., Kärberg, T., Aas, K., Faria, L., Silva, H., Ferreira, M., and Fereira, B. (2016). *Common Specifications for Information Packages*, version 0.17, DILCIS Board. http://dilcis.eu/images/Specifications/CS/Common_Specifications_for_IPs_v10.pdf

Brocks, H., Kranstedt, A., Jäschke, G., & Hemmje, M. (2010). "Modeling context for digital preservation," *Smart Information and Knowledge Management*, E. Szczerbicki & N. T. Ngugen, eds., Springer, pp. 197-226.

doi:10.1007/978-3-642-04584-4

Burbules, N. C., & Warnick, B. R. (2006). "Philosophical inquiry," *Handbook of Complementary Methods in Education Research*, Routledge, pp. 489-502. doi: 10.4324/9780203874769.ch29

Catt, I. E. (2014). "The two sciences of communication in philosophical context," *Review of Communication* 14(3-4): 201-228. doi:10.1080/15358593.2014.986876

Cochrane, E., Rechert, K., Oberhauser, J., Anderson, S., Fox, C., & Gates, E. (2022). "Useable software forever: The emulation as a service infrastructure (EaaSI) program of work," *18th International Conference on Digital Preservation*, Glasgow, pp. 40-52. https://www.dpconline.org/docs/miscellaneous/events/2022-events/2791-ipres-2022-proceedings/file

Condon, P. B. (2014). *Digital Curation Through the Lens of Disciplinarity in the Development of an Emerging Field*, PhD dissertation, Simmons University. https://beatleyweb.simmons.edu/scholar/files/original/c79af2ebd58dfb5cdaf3a26d95f03419.pdf

Conway, M. (2017). *Policy Domains: A State-Machine Based Approach to Policy-Based Data Management*. doi:10.17615/3330-ke07

CCSDS [Consultative Committee for Space Data Systems] (2019). *Reference Model for an Open Archival Information System (OAIS)*, Recommended Practice CCSDS 650.0-P-3, Pink (pre-magenta) book. https://cwe.ccsds.org/moims/docs/MOIMS-DAI/Draft%20Documents/OAIS%20v3/OAIS%20final%20v3%20draft%20with%20changes%20wrt%20OAISv2%2020190924-rl.docx

Dallas, C. (2007). "An agency-oriented approach to digital curation theory and practice," *International Cultural Heritage Informatics Meeting*, Toronto. http://www.archimuse.com/ichim07/papers/dallas/dallas.html

Danermark, B., Ekström, M., Jakobson, L., & Karlsson, J. C. (2002). *Explaining Society: Critical Realism in the Social Sciences*, Routledge, pp. 80-81.

Dappert, A., & Farquhar, A. (2009). "Significance is in the eye of the beholder*," 13th European Conference on Research and Advanced Technology for Digital Libraries*, Corfu. http://www.planets-project.eu/docs/papers/Dappert_SignificantCharacteristics_ECDL2009.pdf

Dickey, W. 1987). "The mill and the store," *New England Review and Bread Loaf Quarterly* 10(1): 98-111. https://www.jstor.org/stable/40241889

DPC [Digital Preservation Coalition] (2022). *Core Requirements for a Digital Preservation System*. https://www.dpconline.org/docs/knowledge-base/2581-core-requirements-for-a-digital-preservation-system-v1

DPC (2015). "File formats and standards," *Digital Preservation Handbook*, ed. N. Beagrie (Glasgow). https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards

Dugan, R. E., & Hernon, P. (2022). "Outcomes assessment: Not synonymous with inputs and outputs*," Journal of Academic Librarianship* 28(6): 376-380. doi:10.1016/S0099-1333(02)00339-7

Duranti, L., & Thibodeau. K. (2006). "The concept of record in interactive, experiential and dynamic environments: The view of InterPARES," Archival Science 6(1): 13-68. doi:10.1007/s10502-006-9021-7

DuraSpace. (2018). *Portland Common Data Model*. https://github.com/duraspace/pcdm/wiki

Flecker, D. (2000). "Harvard's Library Digital Initiative: Building a first generation digital library infrastructure." *D-Lib Magazine* 6(11). https://www.dlib.org/dlib/november00/flecker/11flecker.html

Freed, N., Klensin, J., & Hansen, T. (2013). *Media Type Specifications and Registration Procedures*, IETF BCP 13, RFP 6828. https://www.rfc-editor.org/rfc/rfc6838.html

Gadatsch, A. (2023). "Modeling and analysis of processes," *Business Process Management* (Wiesbaden: Springer), pp. 101-173. doi:10.1007/978-3-658-41584-6_5

Goethals, A., Frey, F., & Ackerman, D. (2015). *IS&T Archiving Conference*, Los Angeles.

Goethals, A., Frey, F., Wendler, R., Vicary, C., & McEwen, S. (2014). *IS&T Archiving Conference*, Berlin.

Goethals, A., & Patterson, T. (2018). "The big migration: Lessons learned at the completion of the 10-year DRS2 project," *15th International Conference on Digital Preservation*, Boston.

Görzig, H., Gernhardt, B., Engel, F., Vogel, T., Moore, R., & Hemmje, M. L. (2016). "Analysing and representing data management dimensions in research and innovation actions," *World Digital Libraries* 9(1): 1-18. doi:10.18329/09757597/2016/9101

Grace, P. J., & Perry, D. J. (2013). "Philosophical inquiry and the goals of nursing: A critical approach for disciplinary knowledge development and action," *Advances in Nursing Studies* 36(2): 64-79. doi:10.1097/ANS.0b013e3182901921

Hedstrom, M., & Lee, C. A. (2002). "Significant properties of digital objects: Definitions, applications, implications," *DLM Forum*, Barcelona pp. 218-227. https://ils.unc.edu/callee/sigprops_dlm2002.pdf

HL/LTS [Harvard Library/Library Technology Services] (2023). *DRS Futures Project*. https://sites.harvard.edu/drs-futures/

Hjørland, B. (2017). "Subject (of documents)," *Knowledge Organization* 44(1): 55-64. http://www.isko.org/cyclo/subject

IANA [Internet Assigned Numbers Authority] (2022). *Character Sets*. https://www.iana.org/assignments/character-sets/character-sets.xhtml

IANA (2023). *Media Types*. https://www.iana.org/assignments/media-types/media-types.xhtml

ICOM [International Committee for Documentation] (2022). *Volume A: Definition of the CIDOC Conceptual Reference Model*, version 7.1,2, Bekiari, C., Bruseker, G., Doerr, M., Ore, C.-E., Stead, S., & Velios, A., eds. https://cidoc-crm.org/sites/default/files/cidoc_crm_version_7.1.2.pdf

ISO [International Organization for Standardization] (2012). ISO 14721, *Space data and information transfer systems – Open archival information system (OAIS) – Reference model*.

Jaillant, L. (2022). "How can we make born-digital and digitized archives more accessible? Identifying obstacles and solutions," *Archival Science* 22(3): 417-436. doi:10.1007/s10502-022-09390-7

Jebreen, I. (2012). "Using inductive approach as research strategy for requirements engineering," *International Journal of Computers and Information Technology* 1(2): 162-173. https://www.ijcit.com/archives/volume1/issue2/Paper010222.pdf

Jisc (2018). *JiscSD / Canonical-data-model*, Github. https://github.com/JiscSD/Canonical-data-model/tree/master/Data-Model

Kahn, R., & Wilensky, R. (2006). "A framework for distributed digital object services." *International Journal of Digital Libraries* 6(2): 115-123. https://doi.org/10.1007/s00799-005-0128-x

Kastenhofer, J. (2015). "The logic of archival authenticity: ISO 15489 and the varieties of forgeries in archives,"

Archives and Manuscripts 43(3): 166-180. doi:10.1080/01576895.2015.1074085

Kirchhoff, A., & Morrissey, S. (2016). "Digital preservation metadata practice for e-journals and e-books*," Digital Preservation Metadata for Practitioners: Implementing PREMIS*, A. Dappert, R. S. Guenther, & S. Peyrard, eds., Springer, pp. 83-97. doi:10.1007/978-3-319-43763-7_7

Kneuper, R. (2017). "Sixty years of software development life cycle models," *IEEE Annals in the History of Computing* 39(3): 41-54. doi:10.1109/MAHC.2017.3481346

Kuhlthau, C. C. (2017). "Information search process (ISP) model," *Encyclopedia of Library and Information Science* 4th ed., J. D. McDonald & M. Levine-Clark, eds., CRC Press, pp. 2232-2238. doi:10.1081/E-ELIS4

Kyrillidou, M. (2002). "From input and output measures to quality and outcome measures, or, from the user in the life of the library to the library in the life of the user," *Journal of Academic Librarianship* 28(1-2): 42-46. doi:10.1016/S0099-1333(01)00299-3

Lanigan, R. L. (2013). "Communicology," *International Encyclopedia of Communication*, W, Donsbach, ed., Blackwell. doi:10.1111/b.9781405131995.2008.x

Lanigan, R. L. (2015). "Philosophy of communicology: 'Discourse which embodies itself is communication'," *Review of Communication* 15(4): 349-358. doi:10.1080/15358593.2015.1102408

Lee, C. A. (2011). "A framework for contextual information in digital collections," *Journal of Documentation* 67(1): 95-143. doi:10.1108/00220411111105470

Library of Congress. (2015). *PREMIS Data Dictionary for Preservation Metadata*, version 3.0. https://www.loc.gov/standards/premis/v3/

Library of Congress. (2022). *METS: An Overview & Tutorial*. https://www.loc.gov/standards/mets/METSOverview.v3_en.html

Lindlar, M., & Tunnat, Y. (2017). "How valid is your validation? A closer look behind the curation of JHOVE," *International Journal of Digital Curation* 12(2): 286-298. doi:10.2218/ijdc.v12i2.578

MacKenzie, C. M., Laskey, J., McCabe, F., Brown, P. F., & Metz, R. (2006). *Reference Model for Service Oriented Architecture 1.0*, OASIS. http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=soa-rm

Mingers, J., & Willcocks, L. (2017). "An integrative semiotic methodology for IS research," *Information and Organization* 27(1): 7-36. doi:10.1016/j.infoandorg.2016.12.001

Moore, R., Rajasekar, A., & Nief, J.-Y. (2011). "Demonstration of policy-guided data preservation using IRODS," *IEEE Interntional Symposium On Policies for Distributed Systems and Networks*, Pisa, pp. 173-174. doi:10.1109/POLICY.2011.17

Morris, C. 1964. *Signification and Significance: A Study on the Relations of Signs and Values*, MIT Press.

Morrissey, S. 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. doi:10.1080/13614576.2014.883935

Nöth, W. 1990. Handbook *of semiotics* . Indiana University Press.

O'Sullivan J., & Tilbury, J. (2021). "Using preservation action registries to automate digital preservation," *Journal of Digital Media Management* 9(3): 240-252. https://www.ingentaconnect.com/content/jsp/jdmm/2021/00000009/00000003/art00007

Ossher, H., & Tarr, P. (2022). "Multi-dimensional separation of concerns and the hyperspace approach," *Software*

Architecture And Component Technology, M. Aksit, ed., Springer, pp. 293-324.

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

Riva, P., Le Bœuf, P., & Žumer, M. (2016). *FRBR-Library Reference Model*. https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr-lrm/frbr-lrm_20160225.pdf

Riva, P., Le Bœuf, P., and Žumer, P. (2017). *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information.* IFLA. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf

Rogala, A., & Bialowas, S. (2017). "Theory of communication," *Communication in Organizational Environments: Functions, Determinants and Areas of Influence*, Palgrave, pp. 1-27.

SAA [Society of American Archivists] (2020). *Dictionary of Archives Terminology*. https://dictionary.archivists.org/

Sacchi, S. (2017. *What Do We Mean by 'Preserving Digital Information'? Towards Sound Conceptual Foundations for Digital Stewardship*," PhD dissertation, University of Illinois. http://hdl.handle.net/2142/78440

Samvera. (2018). *Hydra::Works*. https://github.com/samvera/hydra-works

Savolainen, R. (2019). "Elaborating the sensory and cognitive-affective aspects of information experience," *Journal of Librarianship and Information Science* 52(3): 671–684. doi:10.1177/0961000619871595

Stamper, R. K. 1993). "A semiotic theory of information and information systems," *ICL/University of Newcastle Seminar on 'Information'*. https://assets.cs.ncl.ac.uk/seminars/101.pdf

Stokes, P., Kaye, J., Fripp, D., Burland, T., Davey, T., & McGowen, M. (2019). "Jisc Open Research Hub: Integrate, deposit, discover and preserve," *16th International Conference on Digital Preservation*, Amsterdam. doi:10.17605/OSF.IO/FUH2A

Tallman, N. (2021). "A 21st century technical infrastructure for digital preservation," *Information Technology and Libraries* 40(4). doi:10.6017/ital.v40i4.13355

Thibodeau, K. (2017). "Digital preservation," *Encyclopedia of Archival Science*, L. Duranti and P. Franks, eds., Rowan & Littlefield.

Tsakonas, G., & Papatheodorou, C. (2011). "An ontological representation of the digital library evaluation domain," *Journal of the American Society for Information Science* 62(8): 1577-1593. doi:10.1002/asi.21559

UNESCO [United Nationals Education, Scientific, and Cultural Organization] (2019). *Concept of Digital Preservation*. https://en.unesco.org/themes/information-preservation/digital-heritage/concept-digital-preservation

van Lamsweerde, A. (2008). "Requirements engineering: From craft to discipline," *16th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, pp. 238-259. doi:10.1145/1453101.1453133

Wagner, F., Schmuki, R., Wagner, T., & Wolstenholme. 20006. *Modeling Software with Finite State Machines: A Practical Approach*. Auerbach.

Wiegers, K. E., & Beaty, J. (2013). *Software Requirements*, 3d ed., Redmond: Microsoft.

Wilson, T. C. (2017). "Rethinking digital preservation: Definitions, models, and requirements," *Digital Library Perspectives* 33(2): 128-136. doi: https://doi.org/10.1108/DLP-08-2016-0029