Rethinking Digital Preservation

Conceptual Foundations

Stephen Abrams

Harvard University
United States
stephen.abrams@harvard.edu
ORCiD 0000-0003-2326-6672

Abstract – In support of a multi-year initiative to revitalize its core digital preservation infrastructure, the Harvard Library is engaged in an open-ended exploration of an ideal system solution. The individual components of that ideal cohere into abstract functional and informational reference models, which act as aspirational benchmarks for requirements and subsequent procurement and deployment activities. The models are derived through the logical refinement of a small set of high-level axiomatic principles. These reflect a conceptualization of digital preservation as an inherently communicative enterprise with an ultimate goal of complementing the persistence of authentic digital information objects with that of opportunities for legitimate information experiences.

Keywords – abductive inference, abstract reference model, communicology, finite state machine, information experience

Conference Topics – From Theory to Practice

I. Introduction

The Harvard Library began operation of its Digital Repository Service (DRS) in October 2000. At that time, no viable commercial or open-source products were available. Consequently, it was necessary for the Library to build a novel system in-house [1]. Since then, use of the DRS has grown to hosting over 10.6 million digital objects, 890 million files, and 90 formats, totaling over 2 PB. These materials span all content genres critical to the University's research, teaching, and learning mission as well as its administrative operation. While the DRS technical platform has been maintained and incrementally updated over the past two decades [2][3], 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. To address these concerns, the Library is engaged in a generational modernization known as the DRS Futures project. This effort 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 [4].

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 needs and goals explicitly unfettered by considerations of how they 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. In essence, the Library is looking beyond what the 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 is possible only when rooted in robust philosophical

iPRES 2023

and conceptual foundations.

1. EXPLORATORY APPROACH

The process of planning and deploying any significant socio-technical system naturally progresses through stages of initial ideation and subsequent development or procurement [5]. The transition from the intangible considerations of the former to the specifics of the latter is codified in terms of system requirements. These function variously as a specification for development, an evaluative rubric for procurement, and acceptance criteria for formal project completion. Traditional requirements development is approached *inductively* [6], relying on 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 [7]. However, in order to achieve a higher level of confidence in final requirements, inductive results should be complemented by a parallel *abductive* process deriving requirements from a small axiomatic set of accepted first principles [8].

Andow describes abduction as 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 [9]. 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 [10].

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.

2. PHILOSOPHICAL INQUIRY

The foundational basis for the Futures project emerged through a process of Philosophical Inquiry (PI). PI is a qualitative research method deriving meaning from experience through abductive questioning of fundamental assumptions within a domain of practice to propose new, and better, explanatory structures for that domain [11]. In the Futures context, the inquiry began with questions regarding the fundamental nature of the preservation enterprise. The *Encyclopedia of Archival Science* defines digital preservation as "the processes and controls that enable digital objects to survive over time" [12]. This formulation 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 (with emphasis added): "policies, strategies, and *actions* that ensure *access* to digital content over time" [13]; "act of *maintaining* information, independently Understandable by a Designated Community, and with evidence supporting its Authenticity, over the Long Term" [14]; "series of *managed* activities necessary to ensure continued *access* to digital materials for as long as necessary" [15]; "*processes* aimed at ensuring the continued *accessibility* of digital materials" [16] – emphasizes two points. First, that the primary role of domain agents is an *enabling* one, e.g., acting as strategiz*ers*, maintain*ers*, mana*gers*, 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 [17]. In other words, access is an enabling factor for subsequent use, which remains a distinct phenomenon. While Wilson argues that this distinction may be operationally prudent [18], it can

be conceptually 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 [19]. However, the ability to retrieve a well-managed object is distinct from a subsequent capacity to make productive use of it. The parameters of that usage are concerned with post-managerial experience.

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 underlying goals are fundamentally humanistic in nature. These give preeminence to the role of the information consumer [20] and the communicative outcomes of the consumer/content engagement.

The success of an act of preservation-enabled communication is dependent on its consumer-facing consequence. That is, preservation acts are successful if they result in a pertinent change to the consumer's intellectual, psychological, or physical state that otherwise would not have been known, felt, or performed [21]. As any such success is contingent with respect to time, place, person, and purpose [22], digital preservation inherently operates in an subjective 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.

A communicative perspective of the digital preservation domain makes it susceptible to a *communicological* approach. Communicology is the study of individually-embodied human discourse [23], in distinction to disembodied machine-to-machine information-theoretic communication [24] and socially-embodied mass communication [25]. That discourse is viewed as a system of expressive *signs* whose meaning emerges through contingent interpretation by their consumers individually, institutionally, and culturally-positioned in socio-technical space [26]. A "sign" is a high-level abstraction for any information-laden entity that "stands to somebody for something in some respect or capacity" [27]. Stamper extended the traditional tripartite structure of a sign – syntactic form, semantic content, pragmatic experience [28] – to encompass six aspects pertinent for greater applicability to digital information systems [29] (see Table 1). Abrams proposed a seventh, performic, aspect for pertinence to digital preservation [30]. This recognizes that digital objects must be dynamically and contextually *performed* to be susceptible to analog human perception and cognitive interpretation [31][32].

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

Table 1. Philosophical Foundations of Digital Preservation

CONCERN	Managerial				Communicative		
REFERENT	Information object				Information experience		
FOCUS	Artifactual				Experiential		
ABSTRACTION	Carrier	Message			Performance	Environment	Mind
FUNCTION	Reificatory	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	Enabling means				Enabled ends		
MEASURE	Output				Outcome		
METRIC	Trustworthiness				Success		
EVALUATION	Objective				Subjective		

3. Semantic meaning or psychological affect

These generally align with the FRBR Manifestation, Expression, and Work constructs [33], which constitute an essential progression from the (relatively) concrete to the (relatively) abstract. The full set of seven semiotic dimensions similarly represents a continuum of perspectives on the preservation enterprise from the objective to the subjective, spanning three descriptive categories:

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

This terminology is borrowed from subject cataloging theory [34], but is deployed to indicate the range of afforded descriptive scope. For example, while this paper *is* an Office Open XML document, it also is overtly descriptive *of* the derivation of a conceptual domain model for infrastructure refresh, while also being interpretatively *about* the model's novelty and legitimacy as a complement to prior modeling efforts.

Preservation outputs and outcomes are 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 [35], while an outcome is a qualitatively-assessable benefit of an output [36]. That is, an outcome focuses on the experiential impact or difference an output has on the part of its recipient [37].

An ontic manifestation is *integral* if it is complete and uncorrupted [38]; an empiric encoding is *valid* if it conforms to an authoritative definition [39]; a syntactic expression is *authentic* if it expresses what it purports to express [38]; a semantic meaning is *reliable* if its factual presentation is accurate [38]; 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 [40]; a plaistic context is *relevant* if it is fit for a consumer's intentional or serendipitous purpose [41]; and a pragmatic understanding is *legitimate* if it is meaningful for that purpose [42][43]. Since any given encounter with preserved digital material is dependent on 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 [30].

3. CONCEPTUAL FOUNDATIONS

Content analysis of digital preservation policy determines that 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 content [19]. Since these policies establish the implicit social "contract" underlying the interaction between preservation stakeholders and delegated service-providers, whether internal or external to an institutional program [30], these qualities suggest three defining imperatives for the preservation enterprise:

1. Ensuring the existence of *authentic* information *objects*
2. Supporting modalities of *authoritative* information *access*
3. Affording opportunities for *legitimate* information *experiences*

Authenticity is the quality of an object being what it purports to be; authoritativeness, that of being appropriate and reliable for the purpose at hand; and legitimacy, that 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 intentions and expectations that future preservation outcomes encompass the preserved artifact *itself*; the means to *interact with* and *know about* the artifact; and the experiential *results* of that interaction. The authenticity/legitimacy distinction contrasts objective universality (authentic for all) with subjective contingency (legitimate for one). 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 contingent context.

Efforts to ensure these beneficial outcomes over time is complicated by the ever-increasing number, size, complexity, and diversity of digital content available for preservation attention, as well as the continual – and often disruptive – evolution and transformation of the modalities of desired (re)use. These problematic aspects of long-term stewardship can be ameliorated through a comprehensive

programmatic approach to fundamental preservation concerns [44], which encompass various functional categories:

1. *Predilect* – Decide what you intend
2. *Select* – Appraise what is available
3. *Collect* – Obtain what you select
4. *Introspect* – Know what you obtain
5. *Perfect* – Enrich what you know
6. *Protect* – Steward what you have
7. *Direct* – Control how you steward
8. *Project* – Offer what you control
9. *Connect* – Provide what you offer
10. *Reflect* – Assess what you do

These extend the set of categories previously derived by Abrams [45] to provide explicit consideration of curatorial discretion regarding acquisition (selection) [46]; opportunities to augment the representation [47], description, and understanding [48] of objects, behaviors, and contexts (perfection); and programmatic governance and accountability (direction) [49]. Regarding the previously identified categories, predilection encompasses stakeholder consultation, analysis, and prioritization. Collection remains the most decisive preservation imperative: while proactive stewardship doesn't guarantee success, an absence of that stewardship almost surely guarantees failure. Introspection provides intellectual as well as technical characterization, facilitating targeted workflow development and automation. Protection lies at the artifactual core of the preservation endeavor while projection and connection mediate the experiential. Reflection supports continuous programmatic improvement.

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* [30]. In the preservation context, an affordance is a functional capability available to a human consumer *to do* something meaningful *with* a preserved object [50]. For example, the *property* of (quantitative) fixity *affords* the ability to determine (qualitative) integrity. Similarly, the property of an image's defined colorspace 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 a subjective performance [51]. The meaningfulness of the pragmatic response to such a performance is dependent on various frames-of-reference that contextualize the encounter [52]. These include the contexts of [45]:

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 [53]
3. Prior consumption, indicative of alternative interpretive reception and response
4. Collateral lived-experience and proximate purpose of the contemporary consumer, which establish experiential expectations

While the domain concept of representation information is defined in generic terms [14], in practice it has not encompassed the means to represent, capture, and retain all of these diverse contextual positions [53]. New infrastructural systems should provide explicit support for persistent management of and experiential access to authoritative performative behaviors and relevant contextual reference frames.

4. Emergent Infrastructural Principles

Digital preservation is a complex of people, policies, procedures, as well as systems facilitating technically-mediated, but fundamentally human communication across time [54]. Given that technical infrastructure is inherently ephemeral and needs to be refreshed and re-envisioned periodically [55][56], it is appropriate to assert expansive aspirations for its function and operation during its design phase. While these may not be immediately provisionable, they set a benchmark for incrementally-achievable programmatic goals. For the Futures project, these goals include support for:

1. Any content genre, language, structure, form, number, size, and description
2. Any managerial duration (interim, persistent, or permanent) and eventuality (proactive when possible, reactive when necessary)
3. Any stakeholder competency, purpose, and modality

The first group 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 state of expertise, tooling, and capacity as well as controlling curatorial priority.

The design, implementation, and operation of preservation infrastructure should embrace a number of programmatically-significant qualities:

1. *Transparency* – Open decision-making [57]
2. *Stability* – Available at a time and place and in a manner of user choice
3. *Reliability* – Predictable behavior conforming to documented function [58]
4. *Productivity* – Maximal purposive impact with minimal effort
5. *Affordability* – Maximal service function at minimal total cost [59]
6. *Sustainability* – Longevity with minimal demands on necessary resources [60]
7. *Functionality* – Responsive enhancement for ever-evolving needs

These factors address important social concerns of stakeholder adoption, retention, and accountability. At a technical level, they should be complemented with other architectural principles, including:

1. Separation of concerns [61]
2. Elastic scalability [62]
3. Asynchronous operation [63]
4. API-first [64]
5. Extension through (re)configuration rather than coding

The first two principles suggest an approach of decoupled interoperability through stateless microservices. The third promotes fault tolerance and adaptive error recovery with eventual consistency. The fourth ensures uniformity of function for both human and automated agents, maximizing opportunities for access modality, automation, and ecosystem integration. The final principle facilitates infrastructural sustainability and relevance through functional customization and enhancement without recourse to expensive software updates. This also permits a wider range of institutional roles to participate meaningfully in functional improvement.

Taken together, these socio-technical principles contribute to the Futures project's evolving abstract functional reference model (see Figure 1). This encompasses computational components at five tiers of abstraction:

1. *Console* – Interfaces for human and automated agents
2. *Registry* – Persistent state for content and logging of infrastructural processes
3. *Proctor* – Machine-actionable policies and automated enforcement

Figure 1. Functional reference model
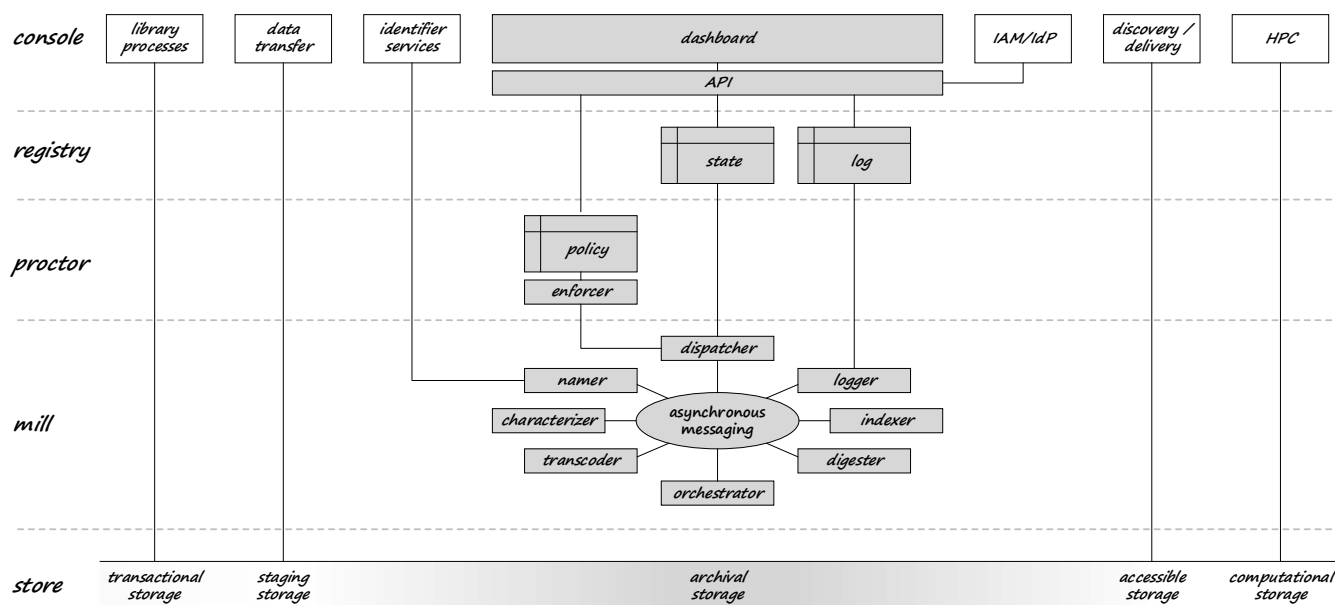
4. *Mill* – Microservice-based processing farm
5. *Store* – Bit-level persistence of tangible manifestations of content (defined by prior Library standardization on the S3 API and OCFL structuring principles [3]).

(The lower two tiers are named in playful homage to Babbage and Lovelace [65].) 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.

The core of the model conceives of ideal digital preservation infrastructure as a finite state machine. 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. An automated policy enforcer evaluates the stimulus in light of current content state and applicable policy rules. If necessary, the enforcer 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 [66][67]. The Preservation Action Registries (PAR) initiative [68] suggests an alternative avenue of exploration regarding the expression and evaluation of policy rules.

Subsequent project activity will supplement these efforts with a stateful information model pertinent to expression and persistence of the artifactual and experiential functions, affordances, and imperatives enumerated in Table 1. The model is still under development, but its current draft form is shown in Figure 2. Its core is a four-level data hierarchy of Objects/Works, Representations/Presentations, Files, and Bitstreams. Works define complex Object aggregations or hierarchies. Objects conform to structural/semantic content *models*, analogous to file-level MIME format typing [69]. This facilitates descriptive high-level characterization and validation, as well as aggregation of like-with-like for efficient bulk processing. Representations (defining subsets of files in the PREMIS sense [70]) and Presentations respectively model static relational file structure and dynamic navigational behavior, similar to the physical/logical distinction of a METS <fileSec> and <structMap> [71]. Similar to Object-level models, Representations are typed by characteristic *tropes* indicating their organizational structure. Files document content independent of specific

instantiations, which are modeled by Replicas, similar to the FRBR Manifestation/Item distinction [33]. The Bitstream entity is introduced primarily to model the heterogeneous contents of container files.

Figure 2. Information reference model
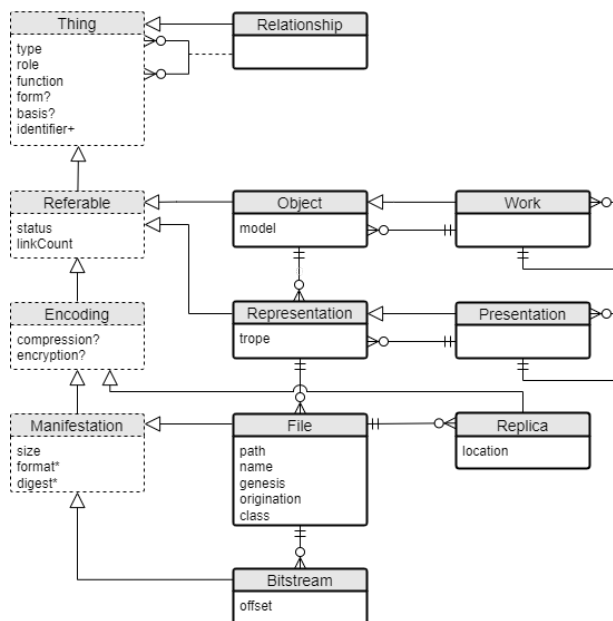
A parallel hierarchy of abstract entities establishes common heritable properties. All are instances of the Thing ur-entity, characterized by their essential type, purposive role, informative function, and expressive form. For example, Objects are of simple or multipart type; Representations, tangible or digital type; and Files, unitary, wrapper, or container type. Similarly, Objects play a (primary) content or (operational) system role; Representations, a substantive, descriptive, or instrumental role; and Files, a data or metadata role. Thing is subtyped to define Referable things and their status – active, (logically) deleted, (physically) purged) – and link count. The latter supports entity composition by reference as well as value. A referable Encoding documents optional compression and encryption as applied to encoded Manifestations representing formatted byte sequences.

5. NEXT STEPS

Once the abstract reference models are fully populated, the generalized use cases and user stories synthesized from the details provided by stakeholder engagement will be aligned with the

derived cases and stories implied by the models. The consolidated cases and stories will inform the development of comprehensive functional and non-functional system and service requirements. These, in turn, will form the basis for a Request-for-Proposal (RFP) to identify plausible candidate solutions. Target candidates will be solicited from commercial vendors and community-supported open-source projects. The RFP also will be evaluated for potential internal Library software development, focusing on the integrative "gluing" together of externally-provisioned components; supplying otherwise unavailable but vital added-value function; or other areas in which the targeted allocation of institutional resources can provide a unique contribution.

6. CONCLUSION

The foundational conceptualization of a domain establishes the metaphoric as well as pragmatic boundaries of legitimate domain focus and action [72]. 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 address 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 enabl*ing* means as well as the enabl*ed* 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 are well-examined and modeled by the broader preservation community at the abstract [14], architectural [73], and deployment [74] levels. Similar efforts regarding the second imperative are emerging through research and practice in software preservation and emulation [75]. Intentions and practices supporting the third, experiential imperative are less mature. The communicological framework proposed here provides useful structuring principles for further investigation of this final preeminent concern.

The Harvard Library DRS Futures project used this communicological framework as the basis for an open-ended exploration of the constituent components of an ideal digital preservation infrastructure. This process derived novel abstract functional and informational reference models from a set of initial axiomatic principles. While the contours of the model infrastructure are unlikely to be fully provisioned in the near term, they nevertheless constitute a critical roadmap for long-term planning of the Library's digital preservation intentions. A future 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 human endeavor, it is very unlikely that achievement ever exceeds aspiration. Thus, there is no reason not to set high aspirations as a benchmark for a desirable goal that can be approached incrementally. The Library hopes that its new conceptual foundation for digital preservation contributes to the success of its internal stewardship priorities, as well as provoking useful community discussion regarding the field's theoretical basis and progress towards state-of-the-art innovation and adoption.

7. REFERENCES

[1] D. Flecker. "Harvard's Library Digital Initiative: Building a first generation digital library infrastructure." *D-Lib Mag.*, vol. 6, no. 11, 2000. https://www.dlib.org/dlib/november00/flecker/11flecker.html

[2] A. Goethals, and T. Patterson. "The big migration: Lessons learned at the completion of the 10-year DRS2 project," presented at the 15th Int. Conf. on Digital Preservation, Boston, 2018.

[3] A. Woods. "A new storage paradigm for sustainable digital stewardship," presented at the CNI Fall 2022 Membership Meeting, Washington, 2022.

[4] Harvard Library and Library Technology Services. *DRS Futures Project*. https://sites.harvard.edu/drs-futures/

[5] R. Kneuper. "Sixty years of software development life cycle models," *IEEE Annals in the Hist. of Comput.*, vol. 39, no. 3, pp. 41-54, 2017, doi:10.1109/MAHC.2017.3481346.

[6] I. Jebreen. "Using inductive approach as research strategy for requirements engineering," *Int. J. of Comp. and Info. Tech.*, vol 1, no. 2, pp. 162-173, 2012. https://www.ijcit.com/archives/volume1/issue2/Paper010222.pdf

[7] K. E. Wiegers, and J. Beaty. *Software Requirements*, 3d ed., Redmond: Microsoft, 2013.

[8] A. van Lamsweerde. "Requirements engineering: From craft to discipline," presented at the 16th ACM SIGSOFT Int. Symp. on Foundations of Softw. Eng., pp. 238-259, Nov. 9-14, 2008, doi:10.1145/1453101.1453133.

[9] J. Andow. "Abduction by philosophers: Reorienting philosophical methodology," *Metaphilosophy*, vol. 47, no.

3, pp. 353-370, 2016, doi:10.1111/meta.12191.

[10] C. M. MacKenzie, J. Laskey, F. McCabe, P. F. Brown, and R. Metz. *Reference Model for Service Oriented Architecture 1.0*, OASIS, 2006. http://www.oasis-open.org/committees/tc_home.php

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

[12] K. Thibodeau. "Digital preservation," in *Encyclopedia of Archival Sci.*, L. Duranti and P. Franks, Eds., Rowan & Littlefield, 2017.

[13] *Definitions of Digital Preservation*, American Library Association, Chicago, 2009. http://www.ala.org/alcts/resources/preserv/defdigpres0408

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

[15] N. Beagrie, Ed. *Digital Preservation Handbook*, 2nd ed., Digital Preservation Coalition, Glasgow, 2015. https://www.dpconline.org/handbook

[16] *Concept of Digital Preservation*, UNESCO, Paris, 2019. https://en.unesco.org/themes/information-preservation/digital-heritage/concept-digital-preservation

[17] *Dictionary of Archives Terminology*, Society of American Archivists, Chicago, 2020. https://dictionary.archivists.org/

[18] T. C. Wilson. "Rethinking digital preservation: Definitions, models, and requirements," *Digital Library Perspectives*, vol. 33, no. 2, pp. 128-136, 2017, doi: https://doi.org/10.1108/DLP-08-2016-0029.

[19] S. Abrams. "Tacit attitudinal principles for evaluating digital preservation success," *Archival Sci.*, vol. 21, no. 3, pp. 295-315, 2021, doi: 10.1007/s10502-021-09360-5.

[20] S. Sacchi. "What do we mean by 'preserving digital information'? Towards sound conceptual foundations for digital stewardship," Ph.D. dissertation, Univ. of Illinois, 2017. http://hdl.handle.net/2142/78440

[21] R. Savolainen. "Elaborating the sensory and cognitive-affective aspects of information experience," *J. of Librarianship and Information Sci.*, vol. 52, no. 3, pp. 671–684, 2019, doi:10.1177/0961000619871595.

[22] B. W. Bishop and C. Hank. "Measuring FAIR principles to inform fitness for use," *Int. J. of Digital Curation*, vol. 13, no. 1, pp. 35-46, 2018, doi:10.2218/ijdc.v13i1.630.

[23] R. L. Lanigan. "Philosophy of communicology: 'Discourse which embodies itself is communication'," *Review of Communication*, vol. 15, no. 4, pp. 349-358, 2015, doi:10.1080/15358593.2015.1102408.

[24] R. L. Lanigan. "Communicology," in *Int. Encyclopedia of Communication*, W, Donsbach, Ed., Blackwell, 2013, doi:10.1111/b.9781405131995.2008.x.

[25] I. E. Catt. "The two sciences of communication in philosophical context," *Review of Communication*, vol. 14, no. 3-4, pp. 201-228, 2014, doi:10.1080/15358593.2014.986876.

[26] J. Mingers and L. WIllcocks. "An integrative semiotic methodology for IS research," *Information and Organization*, vol. 27, no. 1, pp. 7-36, 2017, doi:10.1016/j.infoandorg.2016.12.001.

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

[28] C. Morris. *Signification and Significance*, Cambridge: MIT Press, 1964.

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

[30] S. Abrams. "A communicological critique of evaluative norms for digital preservation success," Ph.D. dissertation, Queensland Univ. of Tech., Brisbane, 2023, doi:10.5204/thesis.eprints.238194.

[31] H. Heslop, S. Davis, and A. Wilson. *An Approach to the Preservation of Digital Records*, National Archives of Australia, 2002. https://www.naa.gov.au/sites/default/files/2020-01/An-Approach-to-the-Preservation-of-Digital-Records.pdf

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

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

[34] B. Hjørland. "Subject (of documents)," *Knowledge Organization*, vol. 44, no. 1, pp. 55-64, 2017. http://www.isko.org/cyclo/subject

[35] R. E. Dugan and P. Hernon. "Outcomes assessment: Not synonymous with inputs and outputs," *J. of Academic Librarianship*, vol. 28, no. 6, pp. 376-380, 2022, doi:10.1016/S0099-1333(02)00339-7.

[36] M. Kyrillidou. "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," *J. of Academic Librarianship*, vol. 28, no. 1-2, pp. 42-46, 2002, doi:10.1016/S0099-1333(01)00299-3.

[37] G. Tsakonas and C. Papatheodorou. "An ontological representation of the digital library evaluation domain," *J. of the American Society for Information Sci.*, vol. 62, no. 8, pp. 1577-1593, 2011, doi:10.1002/asi.21559.

[38] J. Kastenhofer. "The logic of archival authenticity: ISO 15489 and the varieties of forgeries in archives," *Archives and Manuscripts*, vol. 43, no. 3: pp. 166-180, 2015, doi: 10.1080/01576895.2015.1074085.

[39] M. Lindlar and Y. Tunnat, "How valid is your validation? A closer look behind the curation of JHOVE," *Int. J. of Digital Curation*, vol. 12, no. 2, pp. 286-298, 2017, doi: 10.2218/ijdc.v12i2.578.

[40] L. Jaillant. "How can we make born-digital and digitized archives more accessible?" *Archival Sci.*, vol. 22, no. 3, pp. 417-436, 2022, doi:10.1007/s10502-022-09390-7.

[41] C. A. Lee. "A framework for contextual information in digital collections," *J. of Documentation*, vol. 67, no. 1, pp. 95-143, 2011, doi:10.1108/00220411111105470.

[42] C. Dallas. "An agency-oriented approach to digital curation theory and practice," presented at the International Cultural Heritage Informatics Meeting, Toronto, 2007. http://www.archimuse.com/ichim07/papers/dallas/dallas.html

[43] L. Duranti and K. Thibodeau. "The concept of record in interactive, experiential and dynamic environments: The view of InterPARES," *Archival Sci.*, vol. 6, no. 1, pp. 13-68, 2006, doi:10.1007/s10502-006-9021-7.

[44] E. Baucom. "Planning and implementing a sustainable

digital preservation program," *Library Technology Reports*, vol. 55, no. 6, pp. 5-28, 2019, doi:10.5860/ltr.55n6.

[45] S. Abrams. "A foundational framework for digital curation: The Sept domain model," presented at the 12th Int. Conf. on Digital Preservation, Chapel Hill, 2015. https://hdl.handle.net/11353/10.429533

[46] L. Work, N. Tallman, M. Shallcross, and C. Mumma. "The right stuff (over time)," presented at the NDSA Digital Preservation Conf., 2020, doi:10.17605/OSF.IO/875XG.

[47] T. Owens. *The Theory and Craft of Digital Preservation*, Baltimore: Johns Hopkins Univ. Press, 2018.

[48] T. Hutchinson. "Natural language processing and toolsets for archival processing," *Records Manage. J.*, vol. 30, no. 2, 2020, doi:10.1108/RMJ-09-2019-0055.

[49] C. Becker, G. Antunes, J. Barateiro, and R. Vieira. "A capability model for digital preservation," presented at the 8th Int. Conf. on Preservation of Digital Objects, Singapore. https://www.ifs.tuwien.ac.at/~becker/pubs/becker_ipres2011.pdf

[50] M. Hedstrom and C. A. Lee. "Significant properties of digital objects: Definitions, applications, implications," presented at the DLM Forum, Barcelona, 2002, pp. 218-227. https://ils.unc.edu/callee/sigprops_dlm2002.pdf

[51] A. Dappert and A. Farquhar. "Significance is in the eye of the beholder," presented at the 13th Eur. Conf. on Res. and Adv. Tech. for Digital Libraries, Corfu, 2009. http://www.planets-project.eu/docs/papers/Dappert_SignificantCharacteristics_ECDL2009.pdf

[52] M. Bonn, L. Kendall, and J. McDonough. "Preserving intangible heritage: Defining a research agenda," *Proc. Of the Assoc. for Information Sci. and Tech.*, vol. 53, no. 1, 2016, pp. 1-5, doi:10.1002/pra2.2016.14505301009.

[53] H. Brocks, A. Kranstedt, G. Jäschke, and M. Hemmje. "Modeling context for digital preservation," in *Smart Information and Knowledge Management*, e. Szczerbicki and N. T. Ngugen, Eds., Springer, 2010, pp. 197-226, doi:10.1007/978-3-642-04584-4.

[54] S. Abrams. "Theorizing success: A communicological approach to evaluating digital preservation efficacy," *Bull. of IEEE Technical Committee on Digital Libraries*, vol. 15, no. 1, 2019. https://bulletin.jcdl.org/Bulletin/v15n1/papers/abrams.pdf

[55] G. Janée, J. Frew, and T. Moore. "Relay-supporting archives: Requirements and progress," *Int. J. of Digital Curation*, vol. 4, no. 1, 2009, doi:10.2218/ijdc.v4i1.78.

[56] J. Barateiro, G. Antunes, F. Freitas, and J. Borbinha. "Designing digital preservation solutions," *Int. J. of Digital Curation*, vol. 5, no. 1, 2010, doi:10.2218/ijdc.v5i1.140.

[57] D. Lin et al. "The TRUST principles for digital repositories," *Sci. Data*, vol. 7, no. 144, 2020, doi:10.1038/s41597-020-0486-7.

[58] E. M. Corrado. "Repositories, trust, and the CoreTrustSeal," *Technical Services Quarterly*, vol. 36, no. 1, pp. 61-72, 2019, doi:10.1080/07317131.2018.1532055.

[59] N. Tallman and H. Wang. "Seeking sustainability," presented at the 18th Int. Conf. on Digital Preservation, Glasgow, 2022, doi: 10.26207/58rw-kt80.

[60] K. Pendergrass, W. Sampson, T. Walsh, and L. Alagna. "Towards environmentally sustainable digital preservation," *American Archivist*, vol. 82, no. 1, 2019, pp. 165-206: doi:10.17723/0360-9081-82.1.165.

[61] H. Ossher and P. Tarr, "Multi-dimensional separation of concerns and the hyperspace approach," in *Softw. Arch.*

And Component Tech., M. Aksit, Ed., Springer, 2002, pp. 293-324.

[62] D. M. Gerrard, J. E. Mooney, and D. Thompson. "Digital preservation at Big Data scales," *Library Hi Tech*, vol. 36, no. 3, pp. 524-538, 2018, doi:10.1108/LHT-06-2017-0122.

[63] R. Laigner, Y. Zhou, M. A. V. Salles, and Y. Liu. "Data management in microservices," *Proc. Of the VLDB Endowment*, vol. 14, no. 13, pp. 3348-3361, 2021, doi:10.14778/3484224.3484232.

[64] N. Beaulieu, S. M. Dascalu, and E. Hand. "API-first design: A survey of the state of academia and industry," presented at the 19th Int. Conf. on Information Tech. – New Generations, 2022, pp. 73-79, doi:10.1007/978-3-030-97652-1_10.

[65] W. Dickey. "The mill and the store," *New England Review and Bread Loaf Quarterly*, vol. 10, no. 1, pp. 98-111, 1987. https://www.jstor.org/stable/40241889

[66] M. Conway, R. Moore, A. Rajasekar, and J.-Y. Nief. "Demonstration of policy-guided data preservation using IRODS," presented at the IEEE Int. Symp. On Policies for Distributed Systems and Networks, Pisa, 2011, pp. 173-174, doi:10.1109/POLICY.2011.17

[67] M. C. Conway. "Policy domains: A state-machine based approach to policy-based data management," M.S. thesis, Univ. of North Carolina, Chapel Hill, 2017, doi:10.17615/3330-ke07

[68] J. O'Sullivan and J. Tilbury. "Using preservation action registries to automate digital preservation," *J. of Digital Media Manage.*, vol. 9, no. 3, pp. 240-252, 2021. https://www.ingentaconnect.com/content/hsp/jdmm/2021/00000009/00000003/art00007

[69] A. Kirchhoff and S. Morrissey. "Digital preservation metadata practice for e-journals and e-books," in *Digital Preservation Metadata for Practitioners*, A. Dappert, R. S. Guenther, and S. Peyrard, Eds., Springer, 2016, pp. 83-97, doi:10.1007/978-3-319-43763-7_7

[70] P. Caplan. *Understanding PREMIS*, Library of Congress, Washington, 2009; rev. 2021. https://www.loc.gov/standards/premis/understandingPREMIS_english_2021.pdf

[71] *Metadata Encoding & Transmission Standard*. Library of Congress, Washington, 2010. https://www.loc.gov/standards/mets/METSPrimer.pdf

[72] P. B. Condon. "Digital curation through the lens of disciplinarity in the development of an emerging field," Ph.D. dissertation, Simmons Univ., Boston, 2014. https://beatleyweb.simmons.edu/scholar/files/original/c79af2ebd58dfb5cdaf3a26d95f03419.pdf

[73] N. Tallman. "A 21st century technical infrastructure for digital preservation," *Information Tech. and Libraries*, vol. 40, no. 4, 2021, doi:10.6017/ital.v40i4.13355.

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

[75] E. Cochrane, K. Rechert. J. Oberhauser, S. Anderson, C. Fox, and E. Gates. "Useable software forever," presented at the 18th Int. Conf. on Digital Preservation, Glasgow, 2022, pp. 40-52. https://www.dpconline.org/docs/miscellaneous/events/2022-events/2791-ipres-2022-proceedings/file