HARVARD UNIVERSITY
Information Technology

FY23 ITCRB Project Proposal Form for Digital Preservation Repository Service Modernization

February 4, 2022

Contents

Project Value

1. Project Description

Project Name	Digital Preservation Repository Service Modernization
Sponsoring VP, Vice Provost or UCIO	Martha Whitehead, Vice President for the Harvard Library and University Librarian and Roy E. Larsen Librarian for the Faculty of Arts and Sciences
HUIT Sponsor(s)	Stuart Snydman, Associate University Librarian and Managing Director, Library Technology
Document Authors	• Stephen Abrams, Head of Digital Preservation, Harvard Library • Jennifer Jubinville, Manager of Project and Portfolio Management, Library Technology Services • Tricia Patterson, Senior Digital Preservation Specialist, Harvard Library • Stuart Snydman, Associate University Librarian and Managing Director, Library Technology • Vitaly Zakuta, Senior IT Project Manager, Library Technology Services
Functional Area	Harvard Library and Library Technology Services
Mission Alignment	Instructional (Teaching and Learning)
Investment Type	Enhancement/Upgrade
Primary Driver	• Avoid risk • Generate efficiencies from process improvements or simplification • Generate recoverable savings from software, license, or other cost savings

2. Project Vision

Digital Preservation Repository Service Modernization

Vision

Modernize infrastructure and services preserving Harvard's digital scholarly resources and institutional records for future (re)use by faculty, students, administrators, and staff, furthering the University's research, teaching, and learning mission and safeguarding its legacy

Objectives

1. Accept the widest diversity of digital materials for proactive preservation care.
2. Support the widest range of archival and curatorial processing workflows.
3. Integrate smoothly with other Library and University systems.
4. Provide intuitive user interfaces for self-service stakeholder interaction.
5. Develop a cost model that supports sustainable operations and incentivizes adoption.

Measurable Benefits

1. Addition and scalable support of new—and often unanticipated—content types.
2. Comprehensive submission, retrieval, and description via API.
3. Working integrations with key Library and University systems, including but not limited to DASH, Dataverse, and FASRC HPC.
4. Positive focus group response to UI/UX and conformance to WCAG accessibility standards.
5. Coverage of full economic costs while maintaining adoption and use at or above historic levels.

Guiding Principles

- Follow principles of user-centered design
- Design with API-first architecture for scalable, fault-tolerant, asynchronous processing
- Prioritize standards over custom approaches
- Align with Library and University IT goals and strategic priorities
- Consider appropriate commercial as well as open source solutions
- Prioritize adaptive service model responsive to emerging technologies and evolving stakeholder needs, goals, and aspirations

3. Value

From its beginnings in 1638, the Harvard Library has been the guardian of the University's memory. Since then, the Library has flourished as a visionary early-adopter and community leader regarding the successful acquisition, management, preservation, and distribution of scholarly and administrative resources, both tangible and digital. However, maintaining that position – and the benefits accruing from it – requires ongoing reflection, reevaluation, and periodic reinvention. Today's environment of scholarly and administrative activity and communication embraces new and ever-evolving forms of highly dynamic, immersive, and web-based reference collections, teaching materials, research outputs, and organizational records. This necessitates a new**,** forward-looking approach towards the preservation management of the University's digital collections. The new Digital Repository Service (DRS) infrastructure provisioned through this project will strengthen the Library's capability and capacity to ensure the critical **long-term archival accessibility, integrity, authenticity, and usability** of Harvard's existing, expected, and even unanticipated digital materials in the coming years.

This project aligns directly with the third CIO Council Initiative of the Harvard IT Strategic Plan:

> **"3. Digital Content Infrastructure:** In partnership with the Library, *develop the infrastructure to support digital content*, including digital asset management and storage, to improve University-wide ability to access and use Harvard's digital content [emphasis added]."

Assurances regarding access and use must incorporate considerations pertinent not only to a single *point in time*, but also *across* time. Focusing preservation goals on open-ended persistence necessitates a similar perspectival shift in the solution. An approach based on traditional IT backup, which assumes modest durations between backup and retrieval, is insufficient to respond effectively to potential impediments and risks likely to be raised by stewardship over decades or longer. The Library's digital preservation program, on the other hand, incorporates proactive and reactive strategies, procedures, and—as embodied by the DRS—technologies explicitly designed to **counteract the corrosive effects of inevitable technical evolution, obsolescence, and disruption over extended timespans.**

Harvard Library stewardship of the University's digital collections embraces the multiplicity of electronic resources and technologies permeating all aspects of pedagogy, administration, and the broader social sphere in the digital age. Effective and sustainable digital preservation of these materials is central to the success of the Library's mission of ensuring the permanence of the scholarly, institutional, and cultural record for use by current and future generations. For more than 20 years, the Library has successfully relied upon locally developed technical infrastructure for stewarding and protecting the University's rich digital collections. However, the DRS has now **reached the conceptual and operational limits of its current design**. A generational modernization to state-of-the-art infrastructure is **necessary for the Library to continue to fulfill its critical stewardship obligations** regarding today's—and tomorrow's—academic resources and institutional records.

3a. Impact

What academic or administrative problem or goal will this project address?

The continued ability to use Harvard's digital collections and institutional records—as well as reuse them, often in novel and groundbreaking ways—is critical to the success of the University's research, teaching, and learning mission; its smooth administrative operations; and its historic leadership position as a preeminent memory institution protecting the legacy of the broader scholarly and cultural record for future generations. Harvard

Library is widely recognized for the achievement and innovation of its programs for the conservation and preservation of all its collections, both physical and digital. Digital technologies offer an exciting array of new functional capabilities previously unobtainable with physical materials. However, electronic resources are inherently fragile with respect to the ever-on-rushing pace of technological evolution and the inevitable disruption of technological innovation. Without continual proactive preservation management, Harvard's digital collections are at significant risk of impaired usability, irretrievable damage, and loss. While digital preservation is not solely a technical enterprise—it also depends critically upon complementary policies, procedures, and human competencies and decision-making—a robust technical infrastructure remains a necessary foundational component of the Library's digital preservation program. This project addresses the Library's need to modernize its repository infrastructure to keep pace with technological innovation, theoretical and operational advances in the preservation field, and evolving stakeholder needs, goals, and aspirations.

The current Digital Repository Service is the culmination of over 20 years of in-house development, maintenance, and incremental enhancement of a custom system by HUIT Library Technology Services (LTS) and its predecessor, Library Office for Information Systems (OIS). The DRS was initially created with a significant infusion of capital funding to the Library from Central Administration as part of the Library Digital Initiative (LDI) program in 1998-2003. At that time, Library technology was focused primarily on supporting management of physical collections through the online HOLLIS catalog and backend integrated library system. However, the University recognized the potential for the radical transformation of its educational and operational enterprise through the growing availability of digital materials, and funded the Library to develop a comprehensive program of digital stewardship and supporting technology. Of the $12 million allocated to LDI, half was dedicated to content projects and half to a suite of first-generation digital library infrastructure. Besides the DRS, this included a visual resource catalog, archival finding aid catalog, e-reserve system, and the Harvard Geospatial Library. While these all had their independent online identity, they all depended upon the DRS as their central repository for immediate access and long-term preservation. The DRS is still surrounded by a constellation of core Library and University systems and services supporting academic and administrative goals. Initial development of the DRS during LDI consumed approximately $2-3 million of the full capital investment. **A similar degree of institutional commitment is now needed** to revitalize the DRS and ensure its continued success.

The DRS currently manages over 10.5 million catalogued items, 228 million files in 90 formats, and 2 PB of storage. These materials have been contributed by over 60 library, museum, archival, academic, and administrative departments, and encompass institutional records, traditional library and museum materials spanning all academic disciplines, and newer scholarly forms such as electronic theses and dissertations (ETDs) and research data. Contributions to the DRS of *any* digital material of persistent value are accepted from *any* University organizational unit (library, archive, museum, research center, academic department, administrative office, etc.), or Harvard-affiliated individual with explicit organizational sponsorship necessary to provide appropriate curatorial oversight and continuity. That material may be of *any* form, structure, or genre and pertinent to *any* Harvard stakeholder community. Consequently, the range of digital content under proactive DRS management is broad and ever-expanding.

The continual emergence of new forms of scholarly and administrative resources increasingly challenges the limits of the DRS's design and implementation. Examples of modern formats that the current DRS cannot fully accommodate include architectural/CAD designs, forensic disk images, transactional databases, and 3D virtual/augmented reality and other immersive digital formats increasingly found in teaching, research, and administrative contexts. Genre and format diversity also is expected to grow through closer integration with other Library and University content systems such as DASH, Dataverse, and the nascent HUIT Enterprise Content

Management (ECM) system. A generational upgrade to state-of-the-art infrastructure is now necessary to better position the Library to meet the University's ever-evolving digital preservation challenges and opportunities regarding these materials in an effective and sustainable manner.

Describe the solution to this problem.

A digital preservation repository provides safe, reliable, and persistent management of digital resources, ensuring their long-term accessibility, integrity, authenticity, and usability. "Long-term" should be understood as ***any*** period of time sufficient to raise concerns regarding the potential negative impacts of changing technologies and evolving stakeholder communities, needs, goals, and aspirations. The Library's Digital Repository Service supports a set of comprehensive functions for the acquisition, storage, description, enhancement, protection, and retrieval of its managed resources. This project reconceives, designs, selects, and deploys new state-of-the-art infrastructure provisioning these service functions, better positioning the University and the Library for effective and sustainable preservation stewardship.

Thus, the project encompasses three phases of Discovery, Planning, and Implementation. These can be viewed alternatively as first defining what an *ideal* repository could be, then determining what an *achievable* repository should be and the path towards it, and finally, implementing that path with a newly modernized *operational* infrastructure for the DRS. The first phase will incorporate extensive stakeholder engagement to devise a comprehensive picture of stakeholder needs, goals, and aspirations. The second phase derives actionable use cases and prioritized requirements and solicits proposals for infrastructure products—open source, commercial, or some combination—that provide an effective solution addressing the fullest possible set of requirements. The third phase involves the careful coordination of software deployment, acceptance testing, data verification, outreach and training, and the final cutover between the old and new systems and subsequent decommissioning of the old. This entire process needs to proceed without any substantial interruption of production service.

3b. Benefits

Modernizing the repository infrastructure will directly and significantly benefit the over 60 curatorial Library and University departments that currently utilize the DRS for managing their digital collections, as well as the Library and HUIT/LTS staff that support the operation, maintenance, and enhancement of the system. Currently, deposit of new materials is a relatively complex process that unnecessarily complicates acquisition of new content. The onboarding of new departmental units wishing to utilize the repository to preserve their rich digital holdings is also more cumbersome than it should be. Additionally, when a new file format or content genre emerges, Digital Preservation Services and Library Technology Services staff must spend significant resources planning for and developing appropriate support. The new infrastructure will remediate these bottlenecks by simplifying deposit and management of materials within the system, as well as minimizing the time and resources required to support current, anticipated, and emergent digital formats.

By streamlining the acquisition and management of scholarly content in the repository, we will additionally benefit Harvard faculty, students, and scholars by enabling greater access to Harvard's incredible digital collections, therefore facilitating research, teaching, and learning. Harvard Library curators will be able to better manage, preserve, and enable access to and use of Harvard's expanding digital scholarly content as well as increasingly digital institutional records. This latter capacity is critical for interoperation with the HUIT Enterprise Content Management system and support of the University Archive's transition to a "digital-first" records management program by 2026. The new infrastructure will be more adaptive to discovery platforms, allowing faculty, students, scholars, administrators, and staff to have greater access to a wider breadth of Harvard's digital collections.

Additionally, Digital Preservation staff will have increased flexibility for preservation analysis, aiding in intervention for at-risk content and better safeguarding Harvard's legacy.

In addition to the functionality and productivity benefits, the operational advantages of a modernized repository will benefit Library and technology administrators. While the enduring performance of the current repository over the past 21 years is evidence of skillful strategic and technical planning, it has become increasingly challenging to maintain and enhance after decades of use. The current DRS has grown to over 200,000 lines of code, the maintenance of which would be removed with this update. In addition to lowering the maintenance burden, updating the infrastructure will provide greater confidence in the reliability and growth potential of the system in the coming decades. As well, contemporary infrastructures are progressively more cost-efficient and environmentally sustainable, and these requirements will be a cornerstone of the selection and/or design of the new repository. These newly realized economic and environmental benefits will positively affect not only Harvard Library but also the entire University ecosystem.

Measurable Benefits:

Benefit	Who will receive the benefit?	How will you measure the benefit?
Simplified deposit and management of digital content	Library staff (including over 60 curatorial units[i] that currently use the repository as well as additional units that would enjoy a lowered barrier to adoption)	Focus group evaluation, including usability tests; growth rate of content and formats in the DRS
Streamlined and strengthened support for current, anticipated, and emergent materials	Harvard digital collection managers and Library preservation and technology staff	Measured reduction in amount of time required to add support for new content types
Increased access to and use of digital scholarly materials	Faculty, students, staff, scholars	Measured increase in requests to retrieve material from the repository
Improved reliability and economic sustainability	Harvard students, scholars, and staff	Less effort dedicated to operational support by LTS and Harvard Library Digital Preservation; decrease in partial cost-recovery pricing
Improved environmental sustainability	Harvard University and beyond	Measured reduction in energy footprint per gigabyte

Potential Savings:

Type of savings (cost avoidance, recoverable savings)	Who will receive the savings?	How will the savings be realized?
The main savings will be cost avoidance in staff and contractor time for supporting, maintaining, and enhancing the current digital preservation repository infrastructure. Currently we estimate an	HUIT Library Technology Services and Harvard Library	A reduction in the number of staff required to support, maintain, and enhance the service

equivalent of 6.7 FTE needed annually to support, maintain, and enhance the current system.		

3c. Risk of Deferral

What are the most significant risks associated with a substantial deferral or denial of funding for this project?

The longer we defer, the less confident we are in our ability to continue to provide our historic high level of preservation assurance and the less confidence we will inspire in Harvard Library curators as well as scholarly and administrative users and stakeholders regarding the trust they place in the repository to safeguard their valuable and unique collections. Deferral also restricts our ability to respond to evolving new forms of scholarly and administrative content important to the mission of the University in a timely manner. Any deferral of activity also will increase the already high demands placed on HUIT resources to maintain the aging infrastructure, impeding effort on other strategic priorities.

The most pertinent risks of deferral are:

- In the digital age, new and complex content types that require preservation services emerge frequently. The current DRS is not architected to add new content types easily. It is estimated that to add a single, new content type to DRS will require upwards of 0.5 FTE of effort. In modern commercial and open source platforms adding new content types is almost trivial. Deferral therefore creates additional risk of being able to preserve new and dynamic content types.
- We estimate that maintaining and enhancing DRS requires upwards of 6.7 FTE per year. As the codebase ages and new requirements are introduced, this commitment will only increase. While a replacement technology will still require local resources to maintain, they will likely be much smaller. Deferral will therefore consume increasing resources that could be directed to other critical priorities.
- As new innovations continually arise in the field of digital preservation practice, we will not benefit from their incorporation into commercial or community-managed open-source platforms and will have to develop them ourselves at considerable cost.
- Allowing the codebase to age further creates fundamental risks associated with technical debt: steadily increasing risk of new security vulnerabilities, technology dependencies that are no longer supported or maintained, and the introduction of new bugs as functionality is custom built over legacy code.

If this project is not funded or completed in the planned time frame, what workarounds would be needed?

Because of the great benefits of this work and high risk of deferral, the work defined by this project would need to be performed, but this could only occur on a much slower timeline, at the expense of other prioritized needs. During that extended period, the University would be potentially leaving at risk many valuable digital artifacts that currently cannot be supported by the digital repository. Library curators already are accruing a backlog of content that is not appropriately preserved because they are unable to deposit it into the present repository. Additional funding would need to be identified to defray the increasingly high support costs of the current infrastructure. We would also need to implement a holdover storage area to stabilize the content that cannot be ingested into the existing repository.

Can the project be implemented incrementally or on a smaller scale? If so, what would the impact be to costs, schedule and academic/administrative benefits?

This project can be implemented incrementally, but it would end up being more expensive because of the long-term incremental approach, potentially leaving a lot of content at risk during the longer implementation time frame.

4. Schedule

Estimated Start and Finish Dates

Phase	Estimated Start	Estimated Finish
Discovery	07/22	03/23
Planning	04/23	12/23
Implementation	01/24	06/25

5. Project Approach

5a. Overall approach

At a high level, what are the major stages, steps and tasks the project will employ to achieve its goals? (This is intended to be a general overview, and not a project plan.)

The digital preservation system will draw from state-of-the-art archival theory and practice, will fully exploit contemporary technical platforms for scalable computation and storage, and will rely upon community supported open source and/or commercial product solutions.

This project will have a phased approach:

Phase I: Discovery (designing an *ideal* repository): July 2022 through March 2023

- Stakeholder consultation, literature review, concept mapping
- Establish functional and non-functional requirements
- Define independent but interoperable abstract models for storage, processing, data, and interaction as benchmarks for solution evaluation
- Evaluate budget and planned team design to determine whether adjustments are necessary before moving into Planning phase

Phase II: Planning (designing an *achievable* repository): April 2023 through December 2023

- Survey landscape for possible solutions, whether locally developed, open source, or commercial
- Buy versus build decision, followed by an RFP or Agile development plan
- Define migration and QA plan
- Determine support plan
- Make procurement decision
- Evaluate plans for development, migration, and support against budget and team design to determine suitability for Implementation phase

Phase III: Implementation (deploying an *operational* repository): January 2024 through June 2025

- Deploy or develop new solution
- Customize and/or integrate as necessary
- Perform migration of content and/or metadata to new solution
- Perform acceptance testing
- Create documentation and conduct outreach and training
- Release new solution to production
- Decommission legacy system
- Conduct post-delivery assessment

5b. Technical Summary

At a high level, please provide a technical summary of the project.

Fundamentally this project will retire multiple aging applications and replace their functionality with a modern solution that could take one of three forms:

1. A commercial solution that is hosted as Software-as-a-Service;
2. A commercial solution that is hosted on Harvard infrastructure, but is maintained and supported largely by the commercial vendor; or
3. A locally developed solution, hosted on Harvard infrastructure, but assembled using modern and community supported open source and standards-based componentry.

The landscape of potential solutions is relatively well known, although it is possible that additional options may emerge during the investigatory process. Our Discovery phase will lead to a thorough set of evaluation criteria and the Planning phase will apply these for ultimate selection of one of the following products or solutions:

- Commercial digital preservation platforms:
  - Libnova: https://www.libnova.com/
  - Preservica : https://preservica.com/
  - Rosetta, by Ex Libris: https://exlibrisgroup.com/products/rosetta-digital-asset-management-and-preservation/
  - Arkivum: https://arkivum.com/

- Open source digital preservation repository platforms and communities:
  - Archivematica: https://www.archivematica.org/en/
  - Samvera: https://samvera.org/
  - Fedora: https://duraspace.org/fedora/

- Alternative solutions: While the prevailing and conventional approach to implementing a digital preservation system is to deploy a purpose-built commercial or open-source platform, we will also consider alternative approaches. Given the considerable effort we have invested in migrating content to a new storage infrastructure that embeds within it significant digital preservation functionality, we believe there is the possibility of implementing a lean front end and associated tooling for content management and metadata storage. We will therefore investigate the prospect of deploying a commercial or open-source digital asset management system (DAMS) or content management system (CMS) in front of our digital preservation storage layer. This would be somewhat novel for the field, but worthy of consideration.

The functional diagram below illustrates the key components of a modern digital preservation system (see Figure 1). The system elements described here, shown within the shaded area of the diagram, are the components this project seeks to replace and modernize.

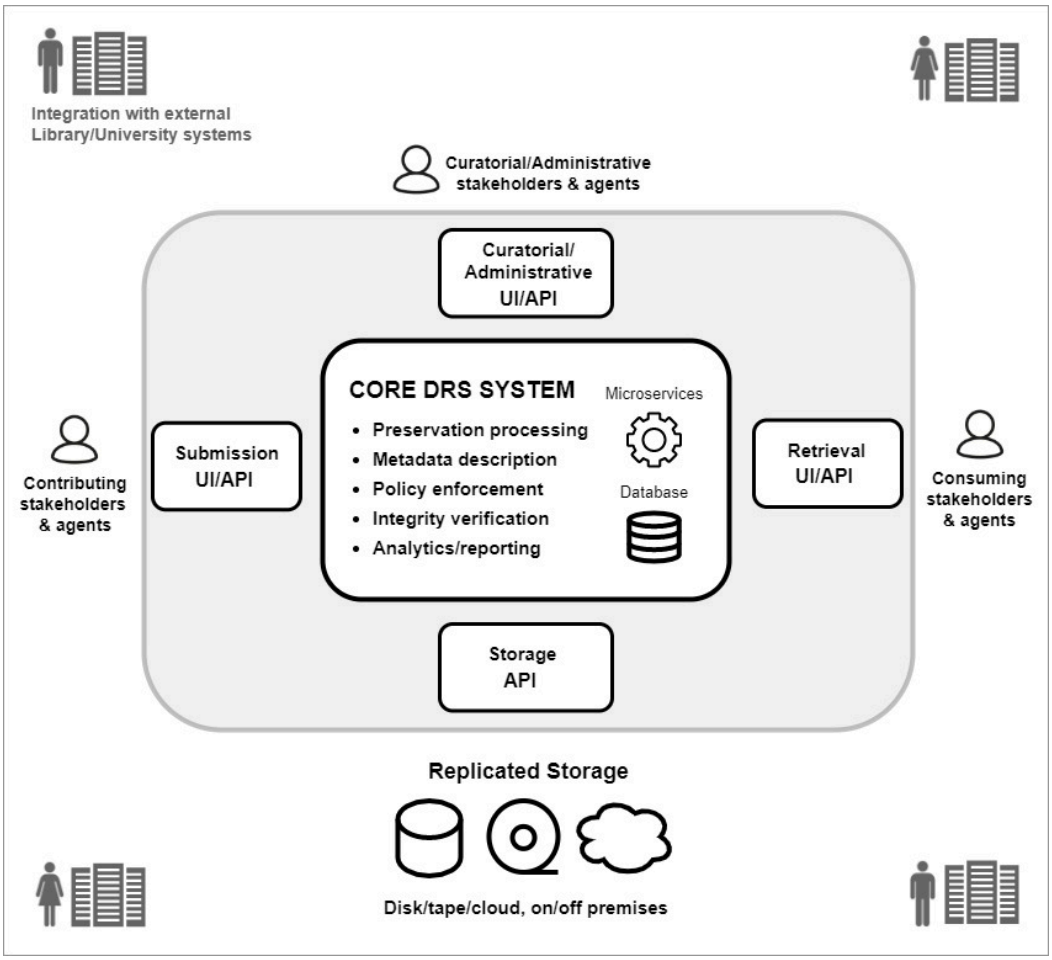

Figure 1 – Functional components of a preservation repository

Core DRS System: The Core DRS System consists of a set of microservices that acts on deposited objects and files to ensure their long-term integrity, accessibility, authenticity, and usability and documents important technical, structural, administrative, and preservation metadata about the content in a database.

Submission UI/API: Individuals and systems can contribute files to the Core DRS System for preservation. A modern web user interface enables the submission of one-off or small batches of files. These types of submissions may come from Library curators, Archives staff, or even faculty who may be enabled to submit publications or datasets. More commonly, content is submitted to the system in large batches via automated pipelines that connect the DRS to other systems. Examples of these pipelines include image, audio, and video digitization labs, electronic theses and dissertation systems, Dataverse, DASH, or Harvard's Electronic Content Management system. A robust submission API enables integration with these other systems.

Curatorial/Administrative UI/API: Once deposited, a web-based dashboard enables curators and content administrators to track the progress of their deposits through the preservation workflow. This dashboard also provides detailed metrics and analytics enabling content administrators to understand and report on their preserved content.

Storage API: Storage replication is a primary preservation strategy employed by the DRS. Each file managed in the DRS is associated with a disposition class assigned by the collection manager exercising curatorial responsibility for the content. That class designation corresponds to a policy regarding the number, location, and media type of the copies made of the given file. For example, content that is curatorially-appraised as of higher value will automatically be subject to greater replication, content intended for archival rather than access purposes will automatically rely more on low-power tape rather than high-power disk, and so on. This scheme permits the appropriate alignment of curatorial intention with technical characteristics, ensuring the most productive use of available storage resources. The storage API is the software layer that enables content to be properly dispositioned. It is carefully designed to allow the easy integration of new storage targets.

Retrieval UI/API: Individuals and systems must be able to retrieve content from the preservation repository to support access and reuse. Retrieval of content from preservation storage may also be necessary in support of disaster recovery. Individuals should have the capability to retrieve content via a web user interface. Systems that enable access and use of preserved content should be able to retrieve files via a robust API.

Replicated Storage: Replicated storage refers to the file storage environments that are used to store replicated copies of the objects to be preserved. Over the course of 2020-2021, LTS has partnered with HUIT and FAS Research Computing to redesign DRS storage. The storage architecture includes on-premises, cloud, and tape media, and leverages storage managed by FAS Research Computing. Part of this redesign was implementing a file storage standard referred to as the "Oxford Common File Layout" (OCFL), which is a community supported open standard specifically designed for long-term digital file preservation. We don't expect major changes to the replicated storage environment as part of this project, and support for OCFL will be an important criterion in evaluating and selecting replacement infrastructure.

6. Project Governance and Stakeholders

6a. Governance

The Harvard Library has a strong existing governance structure to support the Digital Repository Service and related digital preservation activities. Rather than create new governance structures to support this project, we plan to leverage the existing governance structure to provide advice, guidance, support, and oversight to this project. The governance model will include:

- **Executive Committee**: The proposed Executive Committee will consist of senior leadership in HUIT, Harvard Library, FAS, Professional Schools, museums, and administrative departments.
- **Stewardship Standing Committee**: The Stewardship Standing Committee consists of representatives from across the Harvard Library (HL), Harvard College Library (HCL) and the Schools. Its purpose is to steward vulnerable and critical research information in partnership with academic and administrative functions across the University and beyond. Reporting to the Library Leadership Team, the Stewardship Standing Committee shapes, develops, and harmonizes the agenda and action items necessary to steward vulnerable and critical research information in all formats.
- **Born Digital Stewardship Advisory Group**: The Born Digital Stewardship Advisory Group similarly consists of representatives from HL, HCL and the Schools. Its charge is to guide the Harvard Library in developing and implementing strategy for the stewardship of born-digital content managed by Harvard's libraries. This content includes research and licensed data, digital scholarship output, special collections and archival materials, institutional records, course materials, and student and institute publications.

6b. Stakeholders

Who are the primary stakeholders for this project?

Most directly, the primary stakeholders for this project are the 61 curatorial and administrative departments across the Library and University who depend on the DRS for the long-term preservation of their scholarly materials and institutional records. For the most part, however, these departments are working on behalf of the larger constituencies of Harvard's faculty, students, administrators, and staff in support of successful research, teaching, learning, and institutional operations. Harvard's preserved resources are also made available for use by the wider scholarly community.

It is important to note that these stakeholder groups do not exist only at a moment in time. Instead, they are a living chain extending from the ever-forward-moving point of *now* into the indefinite future. Thus, it is not wrong to say that **the ultimate stakeholder for this work is *posterity*:** those future generations of scholars, policy makers, and leaders who will need to draw upon the unique documentary evidence of the past so carefully stewarded for their use through Harvard's institutional commitment, the Library's proactive oversight and continually renewed chain of custody, and technology platforms such as the DRS.

7. Engagement

The Harvard Library has an engaged group of advisors with a significant depth of expertise in all aspects of academic research libraries, their services, and infrastructure. Given the importance of the digital preservation function to the overall mission of the Harvard Library, we will draw on these standing advisory groups for advice, guidance, and feedback.

The Harvard Library **Visiting Committee** consists of experts and Harvard alumni who are appointed by the Corporation. Bi-annual visits provide an opportunity for Visiting Committee members to understand and advise on the Harvard Library's progress towards strategic priorities. Members of the Visiting Committee bring considerable expertise on the state-of-the-art of library services and infrastructure on a global scale.

The Harvard Library **Faculty Advisory Council** consists of Harvard faculty who are deeply engaged and invested in the success of the Library's support of Harvard's teaching and research mission. The FAC meets twice per semester and represents and reflects the interests of the Faculty in their advice on priorities and major investments. The project team will strategically engage with and update the FAC on the progress, milestones, and key decision points of this effort.

Within Harvard, the **Library Leadership Team** consists of the senior leadership of the Harvard Library and College Library as well as the Library Directors of all the schools. The school libraries are primary consumers and beneficiaries of digital preservation services and are dependent upon its success. The Library Leadership team meets biweekly and will be a primary venue for school engagement.

Project Management

8. Project Details

8a. Constraints

What constraints affect this project?

There are a few constraints that affect this project:

- The existing service will need to remain uninterrupted while implementing the refreshed service.
- The costs and availability of digital file storage infrastructure and cloud infrastructure will affect decision-making during the Discovery phase of the project.
- Over 10.5 million objects and 228 million files are already described and cataloged using a range of metadata standards. There is no single all-encompassing preservation metadata standard; instead, the DRS relies on a general preservation metadata standard augmented by special-purpose standards focused on the characteristics of specific content genres and formats. The new infrastructure should accommodate widely accepted metadata standards, and if necessary, provide a smooth transformation pathway from unsupported to supported standards.
- Harvard information security and data privacy policies will be followed.

8b. Assumptions

What assumptions have been made regarding this project (i.e., technical assumptions, use of Harvard resources, availability of services, participation of academic or administration users, etc.)?

The assumptions made regarding this project include:

- Harvard Library Digital Preservation will continue to exercise business ownership of the DRS.
- Management staff from HUIT/LTS and the Harvard Library (specifically, Digital Preservation) will not be backfilled during this project and will be working on this project.
- The project team, governance membership, and allied subject matter experts are committed to the project and its timeline.
- The DRS's current storage infrastructure will be maintained and supported through 2025 and will not need to be revisited as part of this project.
- HUIT and the Harvard Library are jointly accountable for the technical infrastructure that supports the Digital Repository Service.
- The digital preservation technology landscape is such that solutions exist in the open source and commercial marketplaces that can effectively meet our needs and is not expected to dramatically change during the timeframe for the Discovery phase of this project.

8c. Dependencies and Prerequisites

Does the success of this project depend on other projects, initiatives, or services? If yes, please describe. If no, please indicate no prerequisites.

No prerequisites.

Does the success of other projects, initiatives or services depend on the success of this project? If yes, please describe. If no, please indicate no dependencies.

The Harvard Data Commons MVP, funded by the ITCRB Small Ask program, began its work in July 2021 and represents the first step in a broader effort to support the lifecycle of research at Harvard by automating the flow of research data from research computing environments to management, publication, discovery and preservation environments, such as the Digital Repository Service. A resilient, extensible, and performant DRS will be an essential component to ensure the successful evolution of the Data Commons in its next and future phases.

Several of Harvard Library's Multi-Year Goals and Objectives rely on increased functional and technical stability of the DRS:

- Simplify and advance systems to preserve, manage, and access Library digital assets
- Maximize the breadth of tangible and digital collections across Harvard and peer institutions, for the benefit of all partners
- Improve digital infrastructure, in particular to support the preservation of vulnerable audio-visual collections and their use in teaching and learning
- Minimize the environmental impact of collections, services, and spaces
- Simplify and advance systems to preserve, manage, and access Library digital assets

An improved and modernized Digital Repository Service is critical for future interoperation with HUIT's nascent Enterprise Content Management system (a 3-year ITCRB-funded effort) and for support of the University Archives' transition to a digital-only records management program by 2026. The DRS is the venue in which these University records should be deposited, preserved, and stored. A modernized DRS, with the capacity to handle flexible content types and updated user interfaces that facilitate efficient deposit and curation, especially on a larger scale, is required for the University to meet this goal.

8d. Risks

Category	#	Description	Prevention Plan/Action Plan
Schedule	1	Work takes longer than planned, impacting the start of subsequent phases	Granular milestones and frequent check-ins; continual rebalancing and prioritization
Change management	2	Managing user expectations around the adoption of the new environment	Ample consultation and training opportunities, user forums for information sharing
Technical	3	Transition to new system not going smoothly (integration of data models and/or migration of data)	Engage the vendor; add more capacity if needed; make sure there is a strong crosswalk from old to new data model; ensure there is no substantial loss of functions
Financial	4	Potential financial risk with double carrying both environments longer than planned	Slow down planned decreases in pricing to amortize financial obligations

Staffing	5	Tight hiring market could delay filling of term roles	Leave ample time for search; mine existing deep library technology networks; reach out to previous applicants for similar positions; workaround with composite role consisting of several members of the team
Vendor responsiveness	6	Supply chain issues for necessary on-site or hosted equipment	Early procurement to anticipate possible delays; fall back plan for local capacity or cloud resources to undertake the work
Vendor stability	7	Given that this is a multi-year project, and the resulting system is expected to have an extended lifespan, it is important to verify the long-term financial health of the vendor that is being used	Use Harvard Strategic Procurement to verify vendor financial and organizational health; bias potential vendors towards large, well-established companies

9. Change Management

9a. Change Complexity

Actions to implement include:

- Project team will need to develop a change management plan supported by resources
- A Change Management Lead is recommended for moderate complexity projects

The project team will include a person hired specifically for the Change Manager role. Project deliverables and resources will include a change management plan and a comprehensive training strategy and action plan (led by a dedicated Training and Documentation Specialist). The existing DRS service model already includes training protocols; that framework will be beneficial in supporting rollout of change communications and activities as well as training activities to onboard stakeholders and users to the modernized DRS service.

9b. Change Readiness

The project enjoys strong executive leadership support, a mature and engaged governance model and community, and individual stakeholders eager and ready to participate; there is a clear consensus on the need for change and a high degree of stakeholder commitment, as the benefits of this project are well-understood. The project team will leverage these conditions to prioritize stakeholder engagement and carefully execute the change management plan.

9c. Change Resource Needs

Identify how your project will staff the following change management roles:
Change Lead:

The project will hire a Change Manager, who is budgeted to start at the onset of the Planning phase and who will remain on the project team through the conclusion of the Implementation phase. The Change Manager will work in concert with the Project Manager, Business Sponsor, and other key project team members and stakeholders to craft a change management plan and will manage the change management tranche of the effort. Given the expected impact this project will have on curatorial units, the presence of a dedicated change management expert is vital.

Communications Lead:

Considering the DRS's broad stakeholder base and the complexity of the undertaking, prioritizing effective communications will be a cornerstone of the project approach. The Project Manager will take primary responsibility for managing to the communications plan, while Business Owners and School representatives within the project's governance structure, drawing on resources from Harvard Library's communications team, will serve as leads for communication to the broader stakeholder community.

Training Lead:

In recognition that implementation of a revised service will have an impact on the day-to-day work of many Harvard units, the project will hire a Training and Documentation Specialist. This person will join the team in the latter portion of the Planning phase and will stay with the project through the conclusion of the Implementation phase. This person will be charged with developing training and documentation plans and will manage those tranches of the effort, working closely with the Project Manager and other team members.

[i] The following University units and departments contribute to and steward content in the current DRS:

1. Archaeological Exploration of Sardis
2. Arnold Arboretum of Harvard University
3. Arnold Library
4. Baker Library
5. Biblioteca Berenson
6. Cabot Science Library

7. Carpenter Center for Visual Arts
8. Center for Hellenic Studies
9. Countway Library of Medicine
10. David P. Wheatland Collection of Historical Scientific Instruments
11. Davis Center for Russian and Eurasian Studies
12. Dumbarton Oaks Museum
13. Dumbarton Oaks Research Library
14. Dumbarton Oaks Research Library
15. Economic Botany Library of Oakes Ames
16. Eda Kuhn Loeb Music Library
17. Ernst Mayr Library of the Museum of Comparative Zoology
18. Farlow Reference Library of Cryptogamic Botany
19. Fine Arts Library
20. General Artemas Ward Museum
21. Gray Herbarium Library
22. H.C. Fung Library
23. Harvard Art Museums
24. Harvard College Library - Americas, Europe, and Oceania Division
25. Harvard College Library - Environmental Science and Public Policy Archives
26. Harvard College Library - Maps, Media, Data and Government Information
27. Harvard College Library - Widener Library Collection Development Department
28. Harvard College Library - Widener Library Judaica Division
29. Harvard College Library - Widener Library Middle Eastern Division
30. Harvard College Library - Widener Library Slavic Division
31. Harvard College Library - Widener Library Western Languages Division
32. Harvard College Observatory
33. Harvard Divinity School Library
34. Harvard Film Archive
35. Harvard Forest
36. Harvard Graduate School of Design
37. Harvard Kennedy School Library
38. Harvard Law School Library
39. Harvard Library
40. Harvard Map Collection
41. Harvard Planning and Project Management
42. Harvard Semitic Museum

43. Harvard University Archives
44. Harvard University Herbaria
45. Harvard University Herbaria Oakes Ames Orchid Library
46. Harvard University Library Google Project
47. Harvard University Library Office for Information Systems (HUIT Library Technology Services)
48. Harvard-Yenching Library
49. Houghton Library
50. John G. Wolbach Library
51. JSTOR Forum
52. Littauer Library
53. Monroe C. Gutman Library
54. Peabody Museum
55. Radcliffe Archives
56. Reischauer Institute
57. Schlesinger Library on the History of Women in America
58. Tozzer Library
59. Ukrainian Research Institute
60. Ware Collection of Harvard University Herbaria
61. Weissman Preservation Center