Digital Collection Development Policy

Building and Sharing Nationally Recognized Digital Collections

The University of Houston Libraries' Digital Collections are central to the Libraries' vision of becoming a nationally recognized leader among research libraries. In alignment with the University of Houston's Tier One strategic initiative and the University of Houston Libraries' Strategic Plan, Digital Collections support the research and teaching needs of UH faculty and students, fulfill the needs of campus and community constituencies, and collect, present, and preserve digital files of material appropriate for scholarly research.

UH Libraries Digital Repositories

This policy highlights selection criteria for multiple digital repositories managed by UH Libraries in collaboration with outside stakeholders, including the Texas Digital Library and HathiTrust.

Repository	Description
UH Digital Library (UHDL)	The UHDL makes available digitized surrogates of rare and unique materials held by the UH Libraries as well as born-digital items. These materials document the University of Houston, city of Houston, and state of Texas, as well as other historically and culturally significant topics. The repository is managed by UH Libraries.
UH Libraries' Institutional Repository (IR) (Hosted by TDL)	The IR accepts content reflecting the scholarly and administrative output of the university. It makes available digitized and born-digital scholarship and creative works from faculty, students, and staff. The IR also contains digitized and born-digital University publications, . The IR runs on the DSpace platform and is maintained in partnership with the Texas Digital Library.
UH Dataverse (Hosted by TDL)	The UH Dataverse stores and makes available research data from UH faculty, students, and staff. It operates on the Dataverse platform and is maintained in partnership with the Texas Digital Library.
HathiTrust Digital Library	The HathiTrust Digital Library is a large-scale digital preservation repository focused on preserving and providing access to book and journal content digitized from member collections through a number of means, including digitization by Google, the Internet Archive, and through local initiatives. The repository is maintained by HathiTrust.

Content Deposit Locations

Content is deposited into one of the existing digital repositories based on its content type. Note that materials may also live in multiple repositories, based on the functions of those particular repositories. As services and functionality around repositories expand, the defined content areas for any one repository may change or merge over time.

Type of content	Definition	Repositor(ies)
UH Libraries collections/materials	● Digitized versions of analog library collections, including print monographs and serials, archival materials, maps, photographs, audio/visual materials, and musical scores, held by the UH Libraries	Primary: ● UHDL Other: ● HathiTrust
UH faculty, staff, and student scholarship and creative work	● Electronic theses, dissertations, and graduate capstone projects, completed by UH graduate students in fulfillment of degree requirements ● Previously published, peer-reviewed, or juried works—e.g., books, journal articles, music scores, poetry, fiction and creative non-fiction, art, music recordings, and dance—with permission of rights-holder(s) ● Prepublication pre-/post-prints with permission of rights-holder(s) ● Grey literature (e.g., technical reports, white papers, research posters, or symposia proceedings) and other unpublished scholarly works ● "Unpublished" creative work ● Learning objects, including instructional materials and primary sources organized for teaching purposes	Primary: ● IR ● UHDL (for audio, moving image, and visually-oriented digital objects)
Digitized and born-digital University Archives material	● Electronic administrative records produced by university units to document actions and/or decisions ● Student organization digital objects, including rosters, meeting minutes, and	Primary: ● IR ● UHDL (for audio, moving image, and

	event publicity materials ● Web harvesting of website material of enduring value	visually-oriented digital objects)
Datasets	● Raw or processed collections of data, usually presented in a structured form, that are the product of one phase of research and form the basis for further work ○ **Note**: Dataset storage will comply with grant or project guidelines; storage and service provided after the life of the project should be re-evaluated in alignment with library resources.	Primary: ● UH Dataverse

UH Digital Collections Selection Guidelines

Collections and items in the UH Digital Collections should, first and foremost, generate national recognition for the University of Houston and the UH Libraries.

To be included in the Digital Collections, collections and items should also:

● Be of significant research and/or teaching interest
● Align with University of Houston strategic priority areas, especially Energy, Health, or Arts
● Exist nowhere else as digital content that is easily accessible and/or of comparable quality
● Meet existing or anticipated demand

Repository	Criteria for Inclusion

UH Digital Library	● Collections that are worthy of national recognition. They address critical elements including:
	  ○ Supporting research and/or teaching at the University of Houston
	  ○ Aligning with one or more of the University of Houston's strategic priority areas
	  ○ Aligning with popular research trends or issues
	  ○ Making collections of unique or rare items of historical or cultural significance more accessible

	○ Complementing or completing existing digital collections (local or external)
	○ Complementing collections selected for commercial digital publication
	
	● Collections that fulfill important local needs. They address critical elements including, for instance, important local projects and university connections:
	  ○ Supporting a program or project at UH Libraries or the University of Houston (e.g., a conference)
	  ○ Has potential for external funding
	
	Technical criteria (for materials needing digitization):
	
	● Can be digitized without damaging original
	● Has a desirable preservation outcome (e.g., reduced handling of fragile original)
	● Will satisfy requests for digital surrogates
	● Would provide access to otherwise inaccessible content

UH Libraries' IR (Hosted by TDL)	Created by a University of Houston-affiliated individual or unit *Content Prioritization* 1. University publications 2. Reports from administrative bodies 3. College and Departmental publications 4. Faculty scholarship and creative works 5. Theses and dissertations 6. Student scholarship (beyond theses and dissertations) and creative works
UH Dataverse (Hosted by TDL)	The UH Dataverse is a platform for publishing and archiving datasets (and other data products) created by faculty, staff, and students at the University of Houston. Users also have the ability to restrict access to data as desired. ● The UH Dataverse is configured to accept any particular type or subject of research and all file formats are supported regardless of the research type. ● Potential types of research data include: ○ scientific experiments (to include social sciences and

	humanities)
	  ○ input data and simulations results
	  ○ derived data (from processing or combining "raw" or other data)
	  ○ canonical or reference data (gene sequences, chemical structures, etc.)
	  ○ accompanying material, observations, or ephemera
	  ○ oral history recordings
	● The UH Dataverse does NOT accept content that contains confidential or sensitive information, and requires that contributors remove, replace, or redact such information from datasets prior to upload.
	
	● **Notes:**
	  ○ Confidential or sensitive information refers to all identifiable information, such that re-identification of any subjects from the amalgamation of the information available from all of the materials is possible and can include: social security numbers; credit card numbers; medical record numbers; health plan numbers; other account numbers of individuals; or biometric identifiers (fingerprints, retina, voice print, DNA, etc.).
	  ○ The repository defines a dataset as a collection of files, metadata, and ancillary content associated with the data.
	  ○ Formats such as journal articles and conference papers resulting from research should be referred to Institutional Repository. Users have the ability to link related content in different repositories through metadata fields.
	
	Technical criteria:
	
	● File uploads can be up to 2GB per file.
	● Research projects are subject to a 10GB maximum limit.

HathiTrust Digital Library	● Collections that are worthy of national recognition. They address critical elements including:
	  ○ Aligning with popular research trends or issues
	  ○ Aligning with one or more of the University of Houston's strategic priority areas
	  ○ Areas of local emphasis that would offer a unique contribution to the HathiTrust collection

	Technical criteria: ● Previously published in print ● Free of copyright restrictions ● Currently not available elsewhere on the web, including HathiTrust Digital Library, Internet Archive, and Google Books ● Fully described and marked up according to HathiTrust guidelines ● Complies with file format requirements of HathiTrust

Digital Collections Review Process

Focused on creating and implementing policies to increase national recognition around digital initiatives, the Digital Collections Management Committee (DCMC) has the explicit authority to:

- Approve or disapprove proposed projects using selection prioritization policies prior to projects entering any specific digital initiatives workflow
- Determine the repository for projects using digital deposit policies prior to projects entering any specific digital initiatives workflow

Digital collections are selected through the Call for Digital Projects (CFDP) process. During each cycle, DCMC will work closely with the Coordinator of Digital Projects and the Digitization Services Coordinator to prepare the call and communicate selection decisions, including feedback when relevant. The committee will develop a mechanism for distributing the call and proposals will be submitted through an online submission form (http://intranet.lib.uh.edu/dcmc) found on the committee's intranet page. The number of projects selected during each cycle will depend on the extent of the collections and the resources available to digitize, describe, and make collections accessible. If a prior cycle's work has not been completed by the start of a new cycle, DCMC will work with the Coordinator of Digital Projects and the Digitization Services Coordinator to select the appropriate number of collections to include in the new cycle, including the possibility of skipping a call altogether.

In addition to selecting projects during the three cycles, DCMC will also evaluate time-sensitive proposals or those with external funding on a case-by-case basis and will make the appropriate changes to accommodate newly prioritized content. It may also solicit special CFDPs, that may fall outside of the regularly scheduled cycles, when needed.

The selection criteria for the CFDP includes:

A. Collections that are worthy of national recognition. They address critical elements including:

      a. Aligning with one or more of the University of Houston's strategic initiatives, including: Arts, Health, Energy, Athletics.
      b. Making collections of unique or rare items of historical or cultural significance more accessible
      c. Complementing collections selected for commercial digital publication
      d. Completing projects with defined external deadlines or with funding attached to the creation and dissemination of materials

B. Collections that support the UH Division of Research's Research Thrusts
C. Collections that support the research and student success objectives at the University of Houston. They address critical elements including:
      a. Aligning selections with popular research trends or issues
      b. Developing digital projects in conjunction with the Digital Research Center
D. Collections that fulfill important local needs. They address critical elements including:
      a. Important local projects and university connections:
      b. Complementing or completing existing digital collections (local or external)
      c. Has potential for external funding

Reappraisal Guidelines

To better balance the research potential of collections with the necessary allocation of resources (space, staff, time, and conservation resources) for their care and preservation, DCMC should reappraise digital collections over time to determine whether some collections can no longer be supported. This should be an open, structured process conducted in consultation with content selectors and curators.

Policy Review

- Completed by DCMC: June 2, 2015
- Approved by Library Administration:September 23, 2015

Members of DCMC will review and update this document every two years.

Policy Revisions

- 2018-01-26: added information and policies for UH Dataverse, HathiTrust, and the Call for Digital Projects process; revised policies for IR.
- 2015-07-30: revised commercial digital publishers section to state that DCMC reviews any proposal and gives recommendation to library administration. Also defined information that the Head of Special Collections should submit to DCMC with any proposals from a commercial digital publisher.
- 2015-07-28: migrated policy to UH Libraries Digital Initiatives Policies and Guidelines

document. Eliminated "Related Policies and Guidelines" section because it is now part of the DIPG table of contents

References

- Carolina Digital Repository: https://cdr.lib.unc.edu/scontent/aboutPages/policiesGuidelines.xml
- Georgia Southern University: http://digitalcommons.georgiasouthern.edu/collection_development.html
- HathiTrust: http://www.hathitrust.org/take_down_policy
- Salt Lake Community College: http://slccarchivesir.files.wordpress.com/2010/09/ir-collection-development-policy-013009.pdf
- Society of American Archivists: http://www2.archivists.org/sites/all/files/GuidelinesForReappraisalAndDeaccessioning-May2012.pdf
- University of Vermont: http://cdi.uvm.edu/about/colldev