

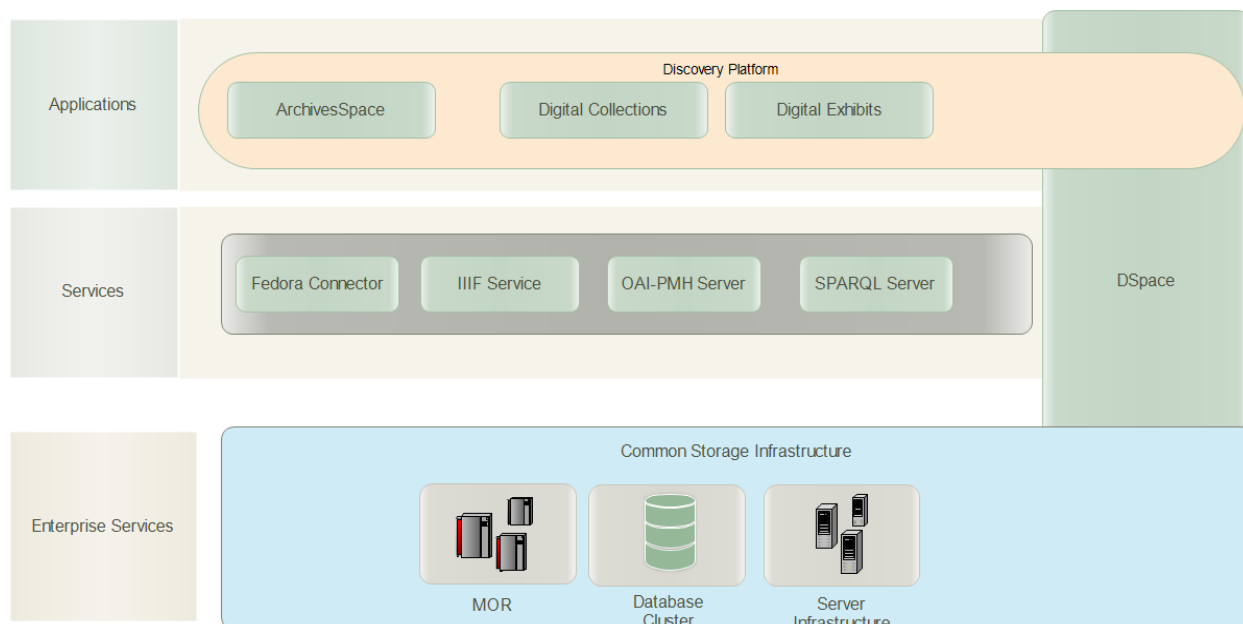
# Digital Content Management Workflow Task Force

## Contents

Digital Content Management Workflow Task Force.....	1
Background .....	1
An Abstract Workflow.....	3
Content Types: .....	5
Preservation by Content Type .....	7
Recommendations: .....	8
Appendix A: Digital Object Decision Tree .....	10
Appendix B: Task Force Membership and Charge .....	11

## Background

Since 2012, the OSU Libraries has been undertaking a thoughtful evolution of the Libraries’ digital collections environment. This process has involved a variety of groups from across the Libraries, and reflects a conscious decision to invest significantly in the development of a robust digital repository infrastructure that can support the Libraries’ growing collections of reformatted and born-digital content. The new digital collections environment seeks to leverage the Libraries’ existing local repository infrastructure (e.g. the Knowledge Bank) and other remote, shared repository infrastructure (e.g. HathiTrust, OhioLINK), in conjunction with the Libraries’ new Fedora-based enterprise infrastructure to provide a robust digital preservation and curation environment capable of supporting the Libraries’ digital collections programs.



OSUL Digital Initiatives Framework

Since the beginning of the Libraries’ digital initiatives and reformatting programs, these efforts have relied heavily on the Libraries’ Institutional Repository (IR), the Knowledge Bank, built upon the DSpace platform, to manage and provide access to many of the digital assets created by the Libraries and the broader campus community. While the IR has served this function admirably for many years and will continue to do so for certain classes of digital content, the burgeoning quantity and variety of digital objects produced and acquired by the Libraries has necessitated a shift in how the Libraries manages and provides access to its own collections. Other internally- and externally-managed platforms, including but not limited to PastPerfect, Media Manager, HathiTrust, the OhioLINK ETD Center, Internet Archive, and Olive Active Paper, have also been employed to manage some silos of digital assets, while others have been stored on Libraries’ servers, staff computers, and on a variety of off-line media.

In order to meet the challenges of effectively managing and preserving this growing diversity of digital object types and creators, the Libraries has moved to implement the new repository infrastructure described in the Master Objects Repository (MOR) Task Force Report<sup>1</sup>. The Libraries’ Applications Development & Support (AD&S) team installed Fedora 4 in late 2014 to serve as the underlying data store and foundation of the MOR infrastructure. Building upon this new platform, the Libraries’ began development of the Image Management System (IMS) to support the curation workflows of image-based content. The IMS marks the first in a set of new applications for curation of digital collections that will be developed upon the new MOR infrastructure to support the curation and preservation of digital collections owned by the Libraries.

As the Libraries moves to implement the new digital collections environment, a new set of guidelines and workflows must be developed to help curators and content managers understand where digital content will reside and how to prepare it for ingest. Likewise, as the Libraries evaluates options for mitigating risk to digital collections through remote disaster recovery and preservation, these workflows will inform the prioritization of investment in these services. To help the Libraries in this effort, this Task

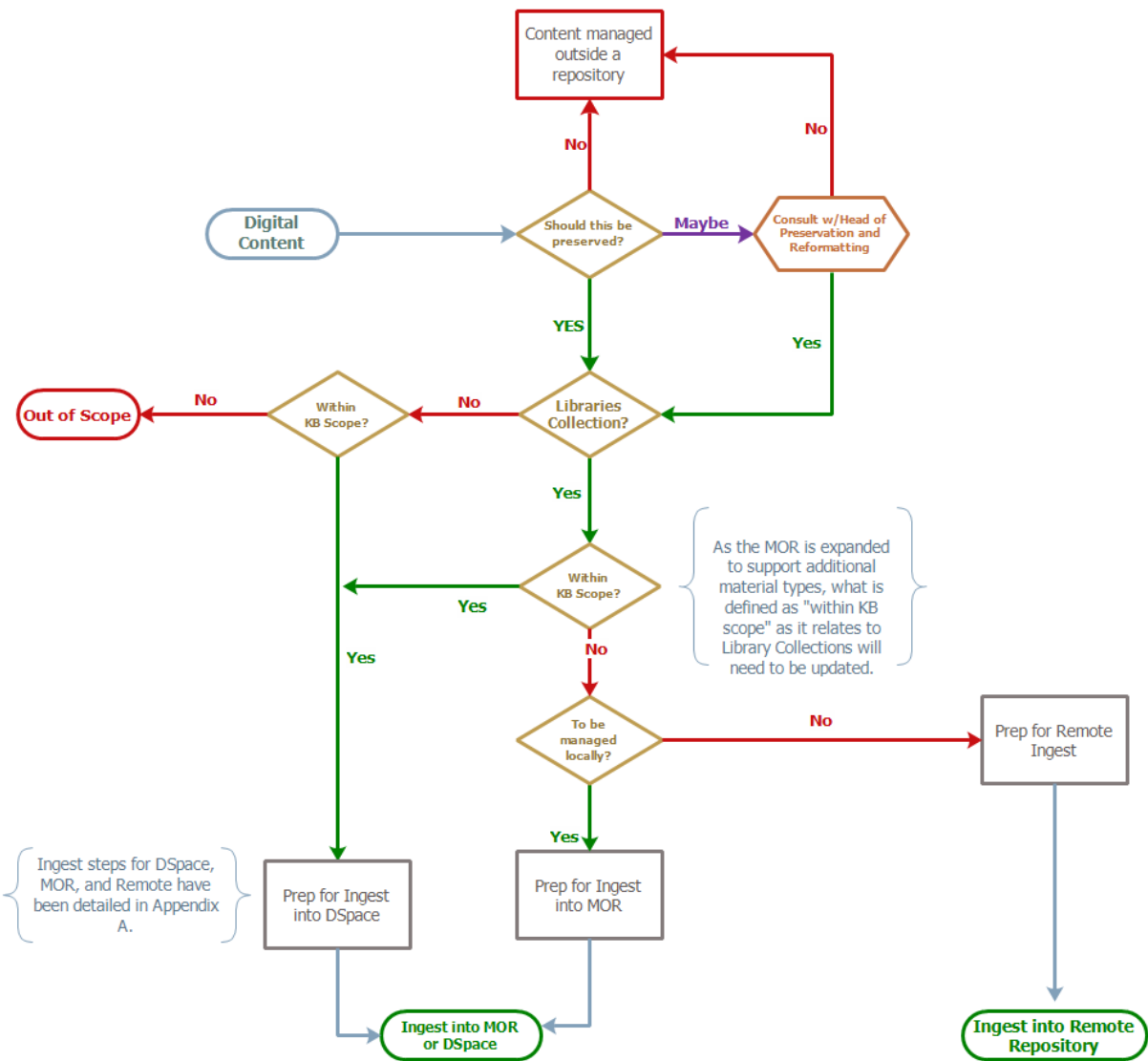
<sup>1</sup> Master Objects Repository Task Force Report. Rev. Nov. 25, 2014. <https://library.osu.edu/document-registry/docs/401>

Force has been charged with developing a set of recommendations detailing how digital assets of various types will move into the Libraries' various digital repositories. The recommendations of the Task Force appear below.

## *An Abstract Workflow*

As the Task Force began to consider how digital assets will move into the Libraries' emerging digital repository infrastructure – including the IR, MOR and other repositories managed locally and remotely – a question emerged: Would it be possible to design a single, abstract workflow model with a common set of decision points and handoffs to describe how any digital content, regardless of type or source, comes into the appropriate repository?

After much deliberation and discussion, the Task Force created a high-level workflow model that can be applied to the vast majority of digital content under the Libraries' stewardship.



*Abstract Digital Content Workflow Diagram*

From this high-level workflow model, the Task Force then began to develop a set of detailed workflows, outlining how common classes of digital assets move into locally and remotely managed components of the Libraries’ digital repository infrastructure. In attempting to create more detailed workflows for common asset classes, a second important question emerged: What is the scope of content that will be managed in each of the local and remote repositories that form the Libraries’ new repository environment? This question is contingent on policy decisions that have not yet been made, and in many cases, workflows that do not yet exist as the MOR framework and supporting curation tools continue to develop. To that end, the Task Force found that it could only develop workflows and an actionable disposition matrix for the most common classes of digital collections materials. Less common asset classes, including those that the Libraries has never actively managed before, will need to be addressed

in the future. Additionally, the Task Force identified specific areas for continued development around local digital collections management, representing recommended workflows for specific content types.

At each step in the abstract workflow, the depositor will need to address a number of questions in order to move to the next step. In considering this abstract workflow and the scope of each repository within the Libraries' evolving preservation infrastructure, the Task Force developed a high-level decision tree (Appendix A) to help guide content curators seeking to manage/preserve content within the Libraries' preservation infrastructure.

## Content Types:

- **Stakeholder Scholarly Content**  
This content isn't defined by a specific media type, but rather spans all digital content types created by the University's community. This content has traditionally been managed through the Libraries' institutional repository infrastructure, and would continue to be so. Digital Content Services has a set of well-defined workflows developed to ensure proper ingest of content and metadata. Content added to the institutional repository is generally going to have some specific characteristics:
  1. The content is open and publicly available; or available following a period of embargo. Historically, one of the Knowledge Bank's most enduring and important legacies related to managing content within the Libraries has been providing space and education around publicly accessible research and scholarship. While many of the Libraries' archival collections require careful management with limited to no direct public access, the institutional repository is currently the Libraries' vehicle for promoting open scholarship and research for the University.
  2. The Content Owner may (most likely will) not be a member of the Libraries. The institutional repository provides preservation for content, but collection owners and curation often occur outside the Libraries.
  3. The content submitted to the institutional repository is the only version of the material that the Libraries will likely have. Since this content is provided by stakeholders outside the Libraries, it may not meet the Libraries rigorous content standards for preservation. For this reason, materials placed within the institutional repository may not meet the Libraries definition of a preservation master, but all will be treated as preservation masters and managed as such to the best of our abilities.
  4. Research Data: As the research landscape continues to shift towards more open data, the Libraries will need to continue to explore, and potentially house, selected research content.
- **OSU theses and dissertations:** This content has historically been managed by OhioLINK and will likely remain at OhioLINK for the foreseeable future. The Libraries is confident in OhioLINK's ability to maintain byte-level preservation of this content due to investments that OhioLINK is making around a long-term, disaster recovery system. While the system is still in negotiation, OSUL participated in the evaluation and selection of the system.
- **Publishing Program Content:** The Libraries' current publications program provides content across a mixed range of platforms, with preservation of content happening primarily on selected materials within DSpace. Within the DSpace community, established workflows exist for moving

content between DSpace and OJS. There are also preservation options available through the Public Knowledge Project, which has created a private LOCKSS network for journals published on OJS. These and other workflows should be explored to find the best fit for our program and our goals.

- Libraries Collections (reformatted or acquired)

The significant investments in the Libraries internal digital infrastructure and preservation capabilities have been specifically focused on providing a more robust preservation and curation environment for Libraries' digital objects and collections. To support these preservation and curation tasks, the Libraries has implemented a Fedora-based infrastructure to provide a better curation environment and separate the storage of a digital object and its metadata from the application that serves the content. As the Libraries has developed the MOR, OSUL has taken the approach of developing custom portals to manage specific digital content types. This iterative approach will enable the Libraries to craft specific workflows based on content type, as development and improvement of the MOR toolset continues. When thinking about Libraries' managed collections, it makes sense to think about this content in terms of object types:

1. Images: Image-based collections are being accessed / managed through the Image Management System. This is the first system developed within the Libraries to support direct ingest and curation of materials into the MOR.
2. Audio and Video Materials: The Libraries has been investigating the use of Avalon to provide a dedicated set of workflow tools around the audio/video encoding/ingest process. *At this point, this workflow needs to be defined and developed.*
3. Institutional Records, Manuscripts, and Papers: Like a traditional institutional repository, the access/management system will need to be able to support "works" objects, i.e. items with multiple digital items attached to a single digital object. Specific use cases for this service would be the management of archival materials like the University President's papers, or donated personal papers. The difference between an institutional records, manuscripts, and papers management system and the IR is that the Libraries will provide controlled and granular access to and management of the content that is added to the institutional records, manuscripts, and papers management system. *At this point, this workflow needs to be defined and developed. This could be supported via the development of a new application interacting with the MOR, as part of the IR via a policy change around open content, or enhancement of future systems.*
4. Digital Exhibits Archives: *At this point, a workflow needs to be defined and developed* for digital exhibits. This will need to be addressed once the digital exhibits program becomes more fully developed.
5. OSUL Monographic digital content: Historically, the IR has included the storage and management of accessible monographic digital content, with master images of pages stored in the "Dark Archive". Rather than develop a system dedicated to monographic materials, the Libraries should look to utilize the HathiTrust to provide access and digital

preservation for monographic content<sup>2</sup>. The HathiTrust is uniquely configured for this purpose—and, once contributed, the content becomes part of a much large digital humanities corpus that can help drive research and improve access to content for all HathiTrust members.

## Preservation by Content Type

The Task Force identified the following specific content types and the workflows that would accompany them:

### Digital Content Disposition matrix:

	Knowledge Bank/ DSpace	MOR: IMS	MOR: TBD	Remote Repository	Notes
Stakeholder scholarly content	✓			✓	The Libraries may recommend deposit of research data in appropriate subject-based community repositories (ex. ICPSR)
OSU Digital Masters Theses / Dissertations				✓ [OhioLINK ETD]	
<i>OSUL Publishing Program Content</i>	✓			?	Long-term preservation of the OSUL publishing program currently is a mixture of systems: OJS and DSpace. In the future, it may make sense to shift to a remote managed process, utilizing the Public Knowledge Project's private LOCKSS network, allowing content to be managed directly from OJS.
OSUL Images		✓			OSUL image content currently resides in both DSpace and the IMS
<i>OSUL Audio/Video</i>			✓		Some A/V content currently resides in DSpace and the ODEE streaming service
<i>OSUL Institutional Records, Manuscripts, and Personal Papers</i>	✓		✓		OSUL currently manages a number of archival documents and OSUL collections such as oral histories (audio/video), sheet music, newsletters, and course bulletins within the IR. Content requiring selective or managed access like the OSU presidential papers, or

<sup>2</sup> Monographic content currently refers to manuscript content like the materials digitized as part of the Google Books project. However, it should be noted that this definition is fluid, and may change to include additional types of monographic content in relationship to how OSUL leverages the HathiTrust.

					donor letters and papers presently are not managed within any repository system.
<i>OSUL Digital Exhibits</i>			✓		OSUL exhibits currently reside in various websites, Omeka and WordPress
<i>OSUL Digital Monographs</i>				✓ <i>[Hathi Trust, Internet Archive]</i>	OSUL currently stores monographic content related to the Google Books digitization project at HathiTrust and the Brittle Books program in Internet Archives. Unpublished monographs, rare books publications, and manuscripts are also found within the IR – but for the purpose of this report, this content is being referenced as Institutional Records, Manuscripts, and Papers.

*Italics = Workflow to be developed*

**Recommendations:**

The Task Force recognizes that no single set of workflows can be created that will meet the needs of all content types and all scenarios, all of the time. The goal of this Task Force was to define processes for management of the most common digital object classes, i.e. the 80/20 concept, and identify workflows for the content that the Libraries will manage day to day. The Task Force recognizes that gray areas will exist, and in those cases, the expectation is that the Digital Reformatting Working Group will work with the Head of Preservation to adjudicate those instances. The Task Force recommends that the Libraries:

1. Articulate a clear delineation between the scope of content managed within the MOR and within the IR, respectively. In developing these workflows, the Task Force recommends that local management of the Libraries’ digital collections be through the MOR, while the IR should be utilized primarily for housing and preserving stakeholder scholarly content. While there will likely be exceptions to this general practice, the Task Force believes that this will help the Libraries establish clear paths for curators to manage Libraries’ content.  
 [FROM EXEC:] . Approved. SDIWG will articulate and internally publicize this policy. They will oversee the initial reassignment of existing collections to the appropriate repository, with DCS and Reformatting WG to determine on-going repository assignments based on the documented policy. Questions regarding on-going assignments will be referred to SDIWG for resolution.
2. Define/develop an Institutional Records, Manuscripts, and Papers Management System for the Libraries’ archival content in these formats.  
 [FROM EXEC:] Approved. Charge a new task force to perform a needs assessment and



requirements gathering through SDIWG with recommendations to come to Exec.

3. Plan and implement a series of information sessions and smaller trainings to present the work of this Task Force, in order to familiarize relevant OSUL faculty and staff with emerging digital content management workflows, address questions and concerns, and define roles and responsibilities around content creation, preparation and ingest.

[FROM EXEC:] Approved. SDIWG to organize/present in coordination with appropriate lead implementers, i.e. the Digital Archivist and appropriate department heads.

4. Actively pursue archiving monographic content<sup>3</sup> within HathiTrust/Internet Archives.

[FROM EXEC:] Approved. Emily Shaw working on arranging content deposit with HathiTrust. She should continue and report back.

5. Work with the OSUL Exhibits Coordinator to identify expectations around preservation of digital exhibit content and structure.<sup>4</sup>

[FROM EXEC:] Approved. Hold for now on any action.

6. Clarify organizational priorities around the Libraries' commitment to reformatting audio and video materials given other organizational priorities and the significant investments needed to support an A/V reformatting program.

[FROM EXEC:] Approved. The Reformatting Working Group will complete an initial collection survey to inform prioritization for reformatting.

Approved. SDIWG will survey Archives and Special Collections curators to identify "born-digital" A/V materials and develop priorities for ingest into the MOR.

7. Investigate workflows for preserving OSUL Publishing Program content, specifically integration opportunities with DSpace or via the Public Knowledge Project for long-term disaster recovery and preservation.

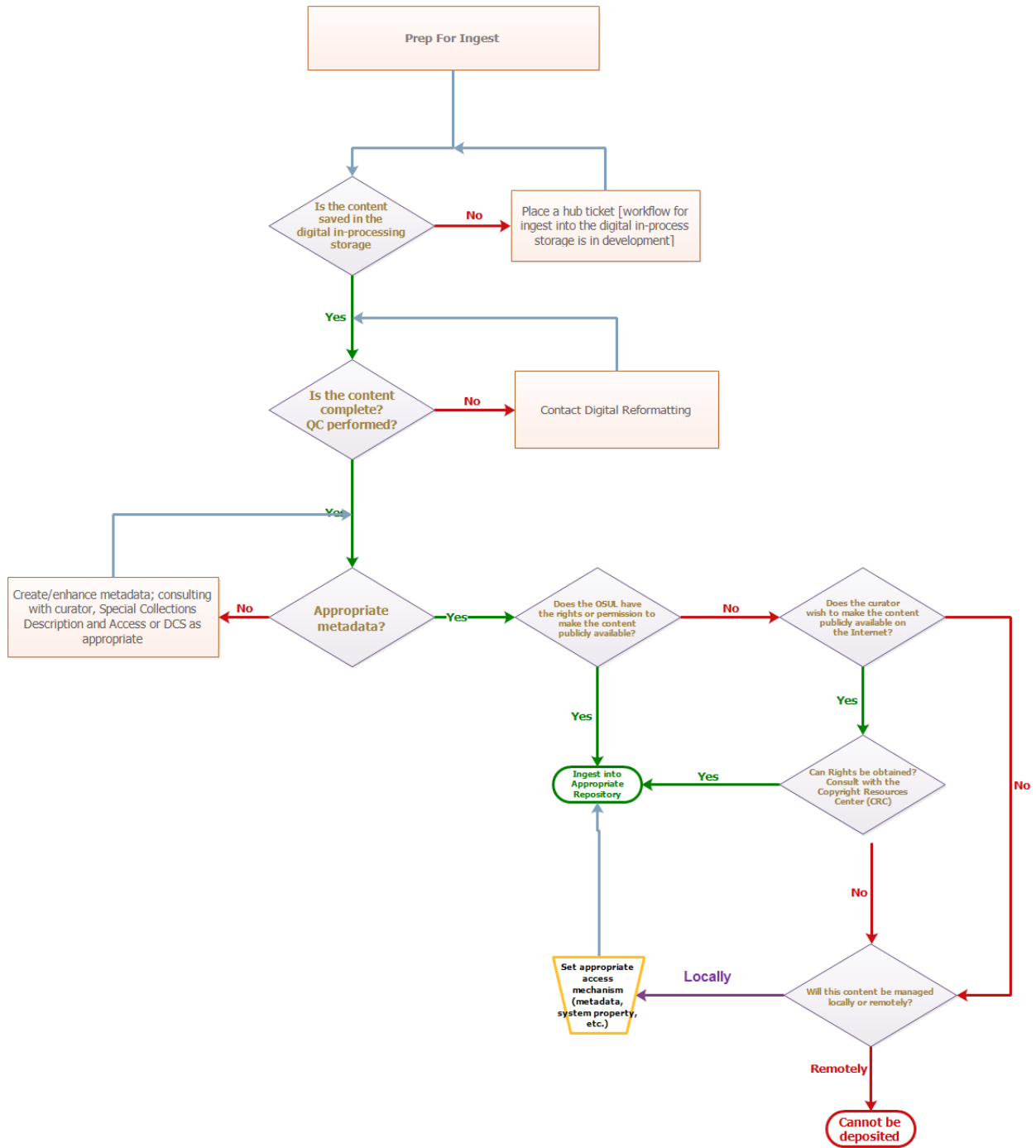
[FROM EXEC:] Approved. DCS will manage in consultation with SDIWG.

---

<sup>3</sup> Monographic content currently refers to manuscript content like the materials digitized as part of the Google Books project. However, it should be noted that this definition is fluid, and may change to include additional types of monographic content in relationship to how OSUL leverages the HathiTrust.

<sup>4</sup> . The Task Force also notes that the Web Archiving Task Force recently submitted a plan for handling archiving web-based content, like digital exhibits.

# Appendix A: Digital Object Decision Tree



## Appendix B: Task Force Membership and Charge

### Digital Content Management Workflow Task Force

**Charge:**

The Digital Content Management Workflow Task Force is charged with developing a disposition matrix and workflow recommendations detailing how digital objects will move into the Libraries' repositories (DSpace and MOR) for management and preservation. This will include reviewing the Libraries' various content systems as well as types of digital content – both reformatted and born digital.

**Strategic Plan Focus Area Supported:**

Focus Area 4.5 of the *Strategic Plan*

**Membership:**

- Terry Reese, convener
- Emily Shaw
- Maureen Walsh
- Melanie Schlosser
- Dan Noonan
- Morag Boyd