This checklist is intended an aid to test for fitness of items to deposit into HathiTrust, or as a general guide for setting up digitization workflows that anticipate deposit into that repository. It is based in large part on the HathiTrust specification for deposit of locally digitized items, but also incorporates other facets of the repository that may affect ingest. It is not appropriate for outsourced digitization, commercial or otherwise. (There is a separate document for those processes.) Each numbered category below is an essential component of the package of each individual work (a "book"). Within each category, components are broken out into thresholds of difficulty beginning with cases of lowest possible barrier (marked as "Ideal") and moving on through cases that may require some remediation representing local effort (marked as "OK"). Using the accompanying spreadsheet will reveal where barriers are low or high at a glance. All questions that arise during evaluation can be referred to the HathiTrust coordinator.

1. All items must be the acceptable in **content type**
    - **Ideal**: Items should be book-like, pages between covers, bound, understood and catalogued as a complete work.
    - **OK**:
        - Items are archival in nature, but are found between book-covers, cataloged as a single work, and have a flow that can be understood as a single, complete work. (ex: a book of bound letters)
        - Structural metadata within a page (for example, denoting the locations of illustrations and captions within a page) will be lost. User experience based on this type of metadata will be lost.
    - **Stop**: If items are other than book-like, they must be discussed with HathiTrust to vet fitness, or determine what might be possible. If user experience is contingent on sub-page-level metadata, items should not be deposited into HathiTrust.
2. ***If*** HathiTrust is expected to be the whole or partial ***delivery*** solution, assure that **criteria for full-viewability** are met. Automated mechanisms in HathiTrust read MARC Bibliographic information and use an automated hierarchy to determine whether to open an item to full view or leave in limited view. This hierarchy can be overridden with appropriate documentation.
    - **Note**: Impact on our yearly fee to HathiTrust will vary based on the decisions in this step. Please coordinate with HathiTrust coordinator to avoid unanticipated rate increases.
    - **Ideal**: Items fall within the criteria that automatically opens them to full view. MARC bib data indicates one of the following
        - that the item is a US government document (will be viewable anywhere in the world)
        - Item is published previous to 1873 (will be viewable anywhere in the world)
        - Item is published between 1873 and 1922, inclusive, and viewability is needed desired/required only within in the United States (will only be viewable only from within the US)
    - **OK**: Items can be excepted through formal mechanisms:
        - Rights holder can submit permissions agreement that will open the items to full view.
        - Cornell Policy Officer has information that can assist HathiTrust manual review of rights information, and successfully open to full-view.

   - **Stop**: Items cannot be guaranteed to meet criteria; delivery from HathiTrust cannot be assured. Note that items can still be deposited into, but not delivered from, the HathiTrust repository.

3. All items must have a reliably unique **item identifier**
   - **Ideal**: items were barcoded at time of digitization, and the barcode was recorded in the item's metadata.
   - **OK**:
     - Barcode can be discovered; work with Lydia Pettis, Joanne Leary or Michelle Paolillo to construct MS Access Query appropriate for discovery.
     - Another identifier exists; work with HathiTrust coordinator to map items to correct namespace in HathiTrust
     - an identifier can be derived from current metadata; work with HathiTrust coordinator to create new namespace and correctly map items on ingest
   - **Stop**: Conditions for assigning unique identifiers cannot be met or are too costly.
4. All items must have **adequate MARC XML** supplied for deposit that conforms to the HathiTrust bibliographic specification. We already have code that produces acceptable MARC XML from the Voyager catalog which can likely be modified to meet needs associated with local submission.
   - **Ideal**: Items were barcoded, cataloged in Voyager and barcode information was recorded in items metadata.
   - **OK**:
     - Items were cataloged and barcodes can be associated
     - Items are housed in a system with metadata that contains MARC XML.
     - Items have some identifier in them (e.g.: OCLC, LCCN, ISSN) that can be used to call another system to reliably produce MARC XML.
   - **Stop**: MARC XML cannot be supplied.
5. All items must be acceptable in **image specification and naming**
   - **Ideal**:
     - **Format:** images containing all page faces sequentially from the outside front cover to the outside back cover.
     - **File format, resolution and color depth**:
       - Pages containing text only, or B&W drawings:
         - 300ppi 10 level gray { png | jp2k | jpg }
         - 600ppi binary { G4 tiff }
       - Pages containing illustrations:
         - 300ppi gray or color as appropriate
         - { png | jp2k | jpg }
       - File specification is noted in file headers.
     - **Naming**: Each image must be named sequentially as found through the book, beginning at the outside front cover: 00000001.tif, 00000002.tif, 00000003.jp2, etc.
   - **OK**:
     - Items do not meet the above guidelines, but can be remediated to meet them. **Note**: if the file headers lack specification, this does not require remediation; this can be handled by the meta.yml file, as described below.

        - Items can be excepted through negotiation with HathiTrust (“Note from Mom”). Refer to HathiTrust coordinator for facilitation of this step.
    - **<u>Stop</u>**: items cannot be remediated or excepted as above.
6. **Each image file** must have accompanying **OCR**
    - **<u>Ideal</u>**:
        - OCR must be Plain text, UTF-8 encoded character set.
        - One OCR text file per image file, and must be named to correspond to appropriate files. (*Example: the OCR file for 00000001.jp2 MUST be named 00000001.txt*)
    - **<u>OK</u>**: OCR is lacking, but can be created.
    - **<u>Stop</u>**: OCR is lacking and cannot be generated.
7. Each item must have an **accompanying meta.yml** file. This file contains resolution and provenance information that is read in the absence of embedded metadata in the image files, preservation information, and structural information that HathiTrust can transform to METS structural data. File is plain text. Format should be after the example at https://docs.google.com/document/d/1iAcgd1zgrVXw3E2enuH6nx_H0qV9wE_XOgQCDxJDuxc/edit
    - **<u>Ideal</u>**:
        - File can be easily derived through existing metadata.
    - **<u>OK</u>**:
        - Required values for file can be discovered through project documentation and supplied.
        - Required values can be reasonably guessed.
    - **<u>Stop</u>**: Required values cannot be discovered or guessed.

The following information is noted for general awareness.

1. **Packaging:** Each item is represented by a ZIP file. Each ZIP file is named after the identifier of the items, and contains
    - No nested folder structure (Use only one container for all files.)
    - An image file for every page of the item, including the inside and outside cover.
    - All OCR text files, one for each image file. (Include blank text file for pages without words.)
    - The **meta.yml** file
    - A plaintext file named **checksum.md5**, which contains a **MD5 checksum manifest** for all other files in the package (image files, OCR files and the meta.yml file).
        - File includes *checksum* (md5) and *filename* for each image and OCR text file and the meta.yml file.
        - Compute the checksums after all remediation and alteration are done; just previous to packaging to assure that each MD5 hash is correct for each file in the package.
2. **Submission:**
    - Packaged items are placed in Box for pickup. Coordinate with the HathiTrust Coordinator for access.
    - MARC XML for all items (one file with multiple records is fine) is submitted to Zephir. Coordinate with the HathiTrust Coordinator for delivery.

  - Ingest reports allow for tracking and remediation of deposit.