Technical Plan - Digitizing the Prohibited Special Collections of The Cannabis Museum

Overview
Over the course of two years, this project will digitize and create descriptive metadata for approximately 10,000 to 15,000 apothecary medical prescriptions, and 400 to 600 pharmaceutical company catalogs / medical journals related to the therapeutic use of cannabis. Cannabis Museum staff will scan or photograph each object, and create item-level descriptive metadata with full text transcriptions. Resulting item data and images will be made freely accessible through the Cannabis Muesum's OMEKA web platform, the Internet Archive and American Institute of the History of Pharmacy's Pharmaceutical Trade Catalog Collection. Prior to award, we will continue to review and update technical plans to match standards used by similar projects and in collaboration with expert organizations will remain open to suggestions.

OMEKA & Plugins Used
The Cannabis Museum has recently moved from the OMEKA.net hosting platform to an internally administered OMEKA instance. OMEKA Classic v2.7.1 is currently installed on our web hosting plan with JustHost.com. We have a PRO Justhost hosting plan, paid out for 20 years, with unlimited data storage and unthrottled bandwidth / no daily visitor limitations. As an open source project, OMEKA has a very active developer community providing modern plug-in architecture for feature development. This includes Batch-Uploader and Dropbox plugins for bulk importing of image files to the OMEKA repository. We use these plug-ins to manage and semi-automate image file upload to our OMEKA install. Exhibit Builder, Geolocation and Exhibit Image Annotation will be used to build collection galleries and annotate relevant pages on prescriptions and catalogs. Other plugins under active consideration include File Paginator, Guest User, History Log, HTML5 Media, Item Relations, Item Review, Neatline, Neatline Widget ~ SIMILE Timeline, Neatline Time, Ngram, PDF Embed, PDF Text, Record Relations, Text Analysis, Text Annotation and UniversalViewer. Key Staff members (Jonathan, Steve) are technically skilled enough to modify plugins and codebase as needed for import and feature development.

Project Equipment - Digitization and NAS
Digital images will be taken utilizing currently available equipment including a Nikon D850 camera with various lenses, EPSON J252A Flatbed Scanner, CZUR-Aura Pro Document Scanner. Images will be created at 600dpi, with JPEG and JP2 derivatives. Project funding will be used to obtain network attached storage (NAS). We are evaluating models such as the BUFFALO TeraStation 3410DN - NAS server - 8 TB, or other market leading equipment. The entire process of photographing or scanning will include cropping, color adjustment, and QA of images using Adobe Photoshop CC. Based on Internet Archive standards, we will use MACAW or similar open-source tool sets for collating and preparing each page image into a complete digital book format.

Preventing Duplicate Digitization
Prior to digitization efforts of pharmaceutical catalogs and medical journal texts, we will review Internet Archive and AIHP's Pharmaceutical Trade Catalog Collection to ensure specific catalogs and issues have not yet been fully digitized. Based on recommendations from AIHP and any repositories discovered prior to the start of this project, we want to ensure digitization efforts are not done in duplication and that items to be digitized are truly unavailable in their full, page-level form. Based on preliminary review, we believe less than 5% of the current Cannabis Museum collection is already fully digitized.

OCR Creation
For the prescriptions, handwritten content will be digitized by subject matter expert key staff (Jonathan, Hannah, Steve) and evaluated by at least one other museum team member prior to making the collection

public. Each digitization of text content will be done to the best and fullest extent possible, denoting any potential conflicts or tentatively understood text.

For the pharmaceutical catalogs and medical texts, individual pages collated into PDF files will be subjected to PDF to Text OCR plugins and community based Scripto OCR libraries. Any issues with OCR plugins will be supplemented by Adobe Acrobat and other open-source OCR scripting libraries to ensure most complete text creation is initially performed. If possible, the project staff will evaluate confidence level data at the page, line, character, and/or word level. Key project staff will also review these initial OCR text creation files, and ensure quality control by random assessment.

Off-site Storage & Cloud Back-Ups
Throughout the project, digital information will be regularly backed-up on portable drives. One of those drives will be maintained off-site. The museum's OMEKA instance and hosted digital files, including those created during this project, are regulatory backed up in AWS cloud storage managed by our web hosting company, and available if needed in the event of an emergency. Project Technical Staff (Jonathan, Steve) will ensure backups will follow a checksum protocol for the detection of possible file corruption.

Workflows and Digitization Standards
All project workflows and metadata standards will align with current local best practices for description and preparation of archival materials for item-level digitization. Several of the museum workflows have been recently evaluated, and will continue to be pre-award. We want to ensure we are following best practices established by information resource and digital humanities organizations. The Cannabis Museum is deeply committed to digitizing its holdings in a manner and way that promotes the widest accessibility by scholars and the public alike. We want to ensure what we do will be available and received by all who are interested in finding and learning from our material. Therefore, prior to the award, this project's PI will remain engaged in evaluating museum workflows and digital tools to affirm that best efforts are made to follow commonly held community standards promoting accessibility and archival-grade digitization.

Current Metadata Elements
Based on the collection item type, the current OMEKA metadata elements are provided below. Pre-award and during initial inventory review prior to digitization these metadata elements will be reviewed and any additional information or fields will be added as needed. We will also ensure standard elements included in online repositories are also included to complete our data digitization effort.

Apothecary Prescription (Rx) Metadata Elements
1. Rx Physician - Doctor who wrote medical prescription, optional gender
2. Rx Patient Name - if available, optional gender
3. Rx Medicine/Drug Compounds - Specific medicine listed on prescription, each drug compound item will be listed separately to improve analysis and data mining efforts. Will start with the number of drugs listed, then automatically open fields to populate. Separate element will ask if "includes Cannabis (Y/N)".
4. Rx Use Instructions - Notes listed by physician on use of the medicine or drug compounds
5. Rx Date - MM/DD/YYYY
6. Rx Number - Text field
7. ID number listed on prescription
8. Rx Originating Address - Street/City/State listed on prescription
9. Notes - open text field
10. Tags - comma separated list
11. Image Files
12. Additional Files - PDFs, Transcription Text files
13. Museum Location - the physical location of the Item (storage binder, location elements)

14. Collection Source - Denotes the source of the collection, e.g.: Wirtshafter, Horowitz, Goldberry
15. Purchase Date
16. Item Unit Purchase Price
17. eBay Listing Title
18. eBay Seller - eBay Identification of Seller Account
19. eBay Seller Notes

Pharmaceutical Catalog Metadata Elements
1. Catalog Title - Title of the Pharmaceutical Catalog
2. Catalog Subtitle
3. Publisher - list of the catalog publisher
4. Published Date - MM/DD/YY
5. Publication Location - City, State
6. Cannabis Relevant Pages - pages where Cannabis is referenced with the book/work.
7. Relevant Pages Notes - Specific passage from the noted Relevant Page
8. Series
9. Volume - Book/Publication Volume
10. Total Pages
11. Dimensions - physical dimensions of the object
12. Format - Print format...e.g.: Hardcover, Paperback, Leather, etc.
13. Conditions Notes - Notes on last inventory schedule & results. Also notes on last time item touched
14. Delivery Date - data object arrived to museum
15. Notes - open text field
16. Tags - comma separated list
17. Image Files
18. Additional Files - PDFs, Transcription Text files
19. Museum Location - the physical location of the Item (storage binder, location elements)
20. Collection Source - Denotes the source of the collection, e.g.: Wirtshafter, Horowitz, Goldberry
21. Purchased From - Person item was purchased from
22. Purchase Date
23. Item Unit Purchase Price
24. eBay Listing Title
25. eBay Seller - eBay Identification of Seller Account
26. eBay Seller Notes

Medical Journals Texts Metadata Elements
1. Journal Title - Title of the Pharmaceutical Catalog
2. Journal Subtitle
3. Publisher - list of the catalog publisher
4. Published Date - MM/DD/YY
5. Publication Location - City, State
6. Cannabis Relevant Articles or Pages - pages where Cannabis is referenced with the book/work.
7. Relevant Pages Notes - Specific passage from the noted Relevant Page
8. Series
9. Volume - Book/Publication Volume
10. Total Pages
11. Dimensions - physical dimensions of the object
12. Format - Print format...e.g.: Hardcover, Paperback, Leather, etc.
13. Conditions Notes - Notes on last inventory schedule & results. Also notes on last time item touched
14. Delivery Date - data object arrived to museum
15. Notes - open text field
16. Tags - comma separated list
17. Image Files
18. Additional Files - PDFs, Transcription Text files
19. Museum Location - the physical location of the Item (storage binder, location elements)

20. Collection Source - Denotes the source of the collection, e.g.: Wirtshafter, Horowitz, Goldberry
21. Purchased From - Person item was purchased from
22. Purchase Date
23. Item Unit Purchase Price
24. eBay Listing Title
25. eBay Seller - eBay Identification of Seller Account
26. eBay Seller Notes