Technical Plan
Peripheral Manuscripts: Digitizing Medieval Manuscript Collections in the Midwest
June 1, 2020 - May 31, 2023
Indiana University, Saint Mary's College, Loyola University Chicago

Indiana University Libraries was an early adopter (2003) of the Fedora digital object repository, which has been used to develop user-oriented services. These services include workflows for back-end processes that are unseen by most users such as image processing, metadata generation, and mapping for metadata sharing (i.e., OAI-PMH), as well as more visible components such as applications for active metadata management leveraged by collection managers and researchers, and public interfaces for discovery.

The *Peripheral Manuscripts* project will use Pages Online at Indiana University, a repository service for preserving and distributing paged, digital image collections held at Indiana University and partner institutions. Pages Online provides standards-based cataloging tools (i.e., enhanced descriptive metadata), workflows (i.e., automatic extraction of MARC metadata through OCLC or local title control numbers) and administrative functions (i.e., ordering pages, creating hierarchical structures, custom labeling, etc.) for optimal curation of paged materials. Pages Online lets collection managers edit and publish digital items in the administrative interface, allowing for ongoing submission and cataloging/structuring of content and the ability to make items available to the public when ready. Viewers of Pages Online interact with high quality, zoomable images following protocols developed by the International Image Interoperability Framework (IIIF). The objects – images and metadata – are stored in our Fedora digital library repository, and can be accessed from our Pages Online web portal: https://pages.dlib.indiana.edu/catalog.

Pages Online is a service developed collaboratively by Indiana University Bloomington and Indiana University Purdue University Indianapolis (IUPUI) for digitized multi-page items with a user-friendly administrative interface. It uses customized Samvera software based on the Plum application developed by Princeton University and the open source Fedora repository software. The Pages Online service launched in 2017 as a pilot for a Samvera- and Fedora 4-based digital object service and repository featuring collection materials from Indiana University Bloomington's Cook Music Library, but will soon expand to include books, newspapers, serial publications, and manuscripts from the IU Libraries' collections and beyond. Efforts are currently underway to rebase its code on the Hyrax software and expand its use cases to include digital image collections. By the time the *Peripheral Manuscripts* is scheduled for its first launch in the Fall of 2021, Pages Online functionality will be far more refined and robust, and should include "unit" (in our case, project-level) and collection level branding for the landing pages. Planned for future iterations of Pages Online, end-users should have the option to download high

resolution images in the public domain. Currently, collection managers which will include partners from this project, will be able to download high resolution images upon request.

This technical plan describes the digitization workflow and the use of Pages Online for curating, organizing, managing and making accessible the manuscripts digitized as part of this project.

Digitization: Capture and Quality Control

All digital image files generated by this project must adhere to format-specific archival practices dictated by "Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files" (http://1.usa.gov/1bqwrlv). With respect to capture, digital files will be created either via a Bookeye 4 overhead scanner or a Hasselblad H1 digital camera using studio setup with copy stand, depending on the condition and characteristics of the materials. When using the Bookeye 4, color images will be scanned at 24-bit color at 600 dpi with embedded Adobe1998RGB color profile. The inclusion of the color profile helps ensure that the image will reproduce as accurately as possible. All master files are scanned at 100% of the original size and saved as uncompressed TIFF files. Color adjustment, when necessary, is done prior to scanning to ensure accurate representation of the original item. The Bookeye will be used for materials that are within size (18" x 24") and do not require special lighting for optimal capture. For codices with tight binding, the Bookeye comes with a v-shaped cradle that accommodates openings between 120 to 180 degrees. The Bookeye also ensures a touch-free digitization process for materials that do not want to lay open. A top glass plate that is self-suspending keeps materials open without undue weight or pressure on the items.

Materials that are candidates for photography include oversize items or manuscripts that contain metallic illumination that will not capture accurately with a fixed position lightsource. The Hasselblad H1 camera is capable of capturing up to 7246 x 5444 pixels, which is well within archival expectations. Mobile lighting is necessary to reduce glare and shadow on items. Variable lighting allows for adjusting the light angle that falls on the object to reduce reflections, shadows, or other problems that can occur. LED lighting is a cool source light that will not heat up and have possible damaging effects on the items like traditional quartz halogen or incandescent bulbs. Filters with studio lighting is a common accessory to correct color temperature, evenness, and overall light balance for accurate capture.

All master files must pass a two-step quality assurance procedure. The first step is an automatic quality control (QC) process that ensures valid and well-formed files. A set of computer programs systematically examines the embedded TIFF tag of every digital file to verify that all files are named according to convention, that they are uncompressed TIFF files, that each file has an embedded profile appropriate to its bit depth, and that all images were scanned at the appropriate resolution. Files that pass automatic QC will be ready for visual inspection. Files that fail automatic QC, will generate a detailed report to aid in correction.

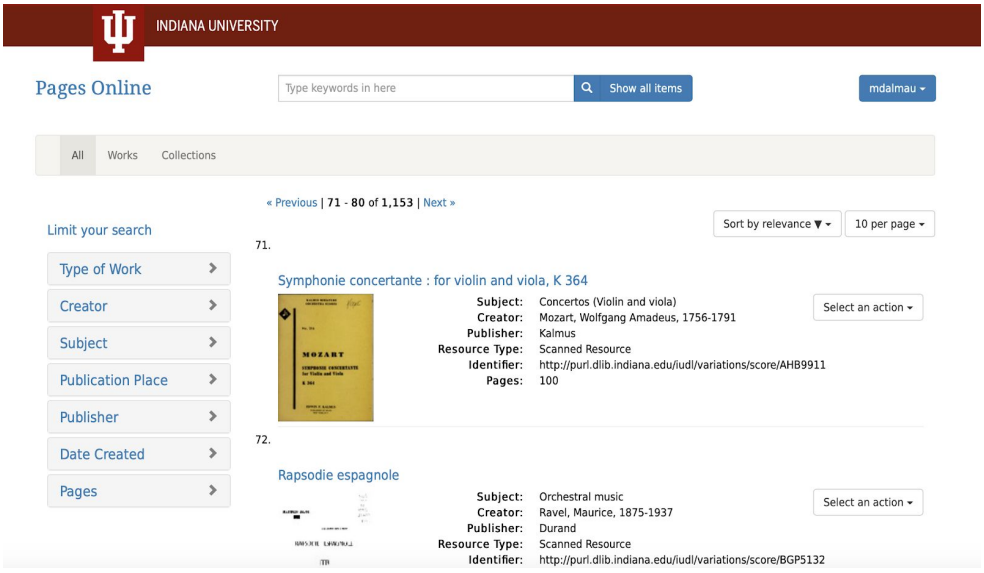
Once the files pass the automatic quality control checks, all of the files are manually examined to ensure accurate visual quality. Each file selected is examined at 100% (1:1) magnification to ensure that the image orientation is correct, that the color balance matches the original items as closely as possible, that it is a sharp, in-focus scan, and that no digital artifacts of scanning are visually present. To aid in the manual quality control process, the physical item is compared with the digital item with attention to color fidelity. If any inconsistencies are found in either the automatic or manual quality assurance checks, the item in question is rescanned until it is acceptable.

Once visual QC is complete, the TIFF files will be staged for ingest to Pages Online. Once the files are submitted to Pages, the Pages application will generate a JPEG2000 for interactive viewing following the J2K Codec specifications, which is a minimum of 400 dpi/4,000 pixels on the long side and a PDF of each manuscript for downloading.

Technical and administrative metadata information will be recorded at scan/capture time following an inventory created by the partners and the PIs as part of their site visits.

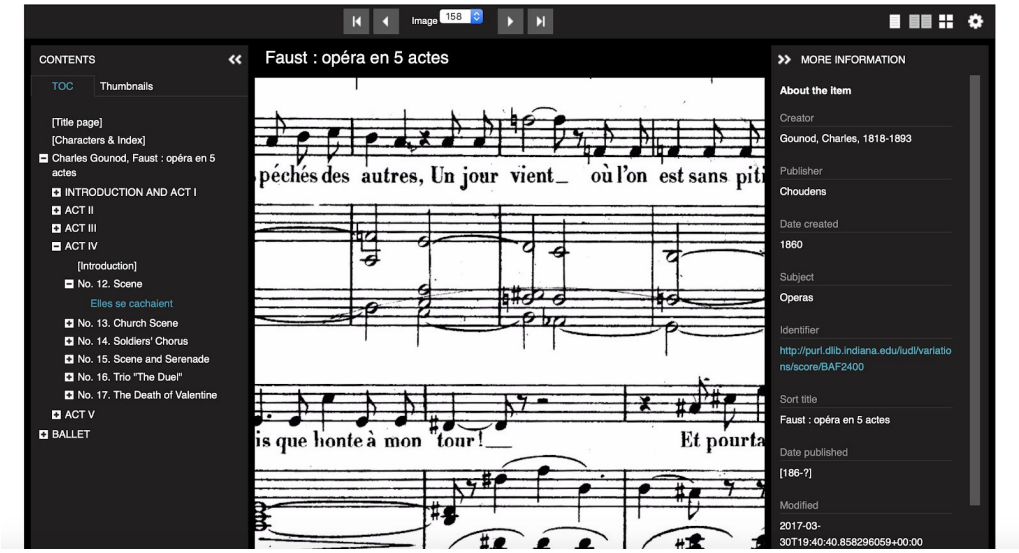
Pages Online: Curation and Access

Currently, Pages Online contains over 1,300 musical scores with approximately 100 openly available to the general public. Although the materials for the *Peripheral Manuscript* project will be open-access, Pages Online provides access controls for content protected by copyright that is being used in the classroom for teaching and learning or other fair use designations. Pages Online uses the open source Blacklight technology to support robust discovery interfaces for content including faceting.

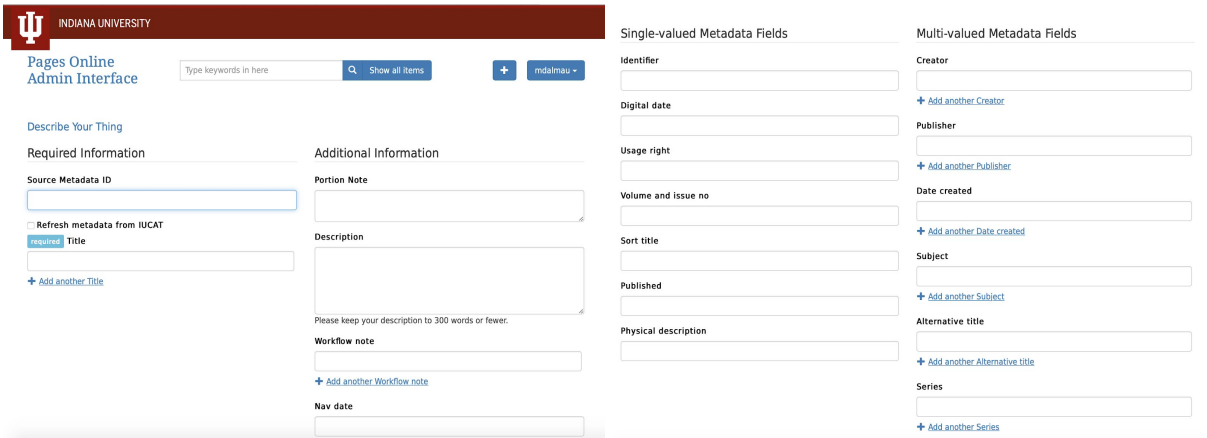


As part of IU's 2020 Bicentennial, Pages Online is committed to ongoing development in order to showcase collections across IU campuses.

Pages Online provides high quality, zoomable images via IIIF Viewer. The full record view provides structural metadata in the form of a table of contents, descriptive metadata including a persistent URL for the item as part of the Pages application and the IIIF manifest, different views -- thumbnails only or two-page layout-- and PDF downloads of the materials.



Current collection managers that use Pages Online retrieve MARC metadata from the IU Libraries online catalog, IUCAT, through supplying a unique identifier in the application. Additionally, descriptive metadata can be adjusted within the 30+ available fields, many repeatable, in the application. Upcoming development for the service will add a flexible metadata feature, which will allow collections to determine their own standards-based metadata models within the service. The metadata model for the project will be determined based on PI site visits and Pages capabilities. The model will be reviewed and solidified as part of the all-partner kick-off meeting.



(The screenshots above only show a subset of the default metadata options. Other administrative operations such as defining color or grayscale PDFs, publishing, etc. are not shown).