Fédération Internationale
des Archives du Film
International Federation
of Film Archives
Federación Internacional
de Archivos Fílmicos

Some Considerations when Setting Up a Digitization Workflow

Anne Gant, Restoration and Digitization & Giovanna Fossati, Head Curator
EYE Film Institute Netherlands

Depending on the scale of the project, each of these steps of a digitization workflow could be relatively simple, or a major undertaking (both in time and budget). For example, a plan could include the following parts: Project planning and workflow design, selecting titles for digitization, selecting appropriate material for digitization, test scanning a variety of material to determine scanning parameters, prepping and cleaning material, moving films physically to and from the scanning location, digitizing the films, verifying and checking digitized raw data, creating metadata about the new digital objects and metadata bout the scanning workflow, moving and storing data, making renditions, verifying renditions, storage and distribution of digital assets, workflow development for use and access of digital assets, documentation and promotion of project, and finally, reporting on the results.

Although every project will differ in budget, goals and strategy, here are some general tips which may be helpful to consider when setting up a workflow.

Plan for technology changes

Especially in a multi-year project, it is possible that technology will advance during the course of the project, and the workflows can be updated. For example, in the next few years as data storage becomes cheaper and data transfer becomes faster, this may affect planning in a positive way. This is wonderful, because the data can be richer, but also can be a problem as a single project will return a wide variety of standards and formats- all with their own migration and playback requirements.

Know what digital asset you want, and what you want to use it for

Is the scan for preservation or access? This might determine the source material used. It might be better to choose an element suitable for preservation (original negative, stored as uncompressed, uncorrected DPXes, deep storage for later restoration) and perhaps choose a different element for a quick-access scan (graded print, color corrected, cropped, with playable renditions in a contemporary aspect ratio).

Create a workflow database or spreadsheet

As projects grow, so does the amount of logistics and workflow information. This might include keeping track of all the scans, re-scans, re-edits, technical questions, requests

and who-did-what surrounding each film. It doesn't have to be complicated; we started with Excel and ended up simply with a FileMaker database. It could include all the steps of the workflow, possible statuses in each step, and the possibility to loop back and re-do any part of the workflow again.

Plan enough preparation time for the material

Analyzing the best film elements for scanning can take time, especially if you have lots of elements for a single film. Do they need to be cleaned, repaired? Do they have enough leader to go through the scanner? Are the elements clearly marked for the people who are scanning? Plan time to add them to the tracking database, or document the scanning process.

What about sound scanning?

If the scanner is not simultaneously scanning sound and matching it to the image, then it will be necessary to sync sound together later. If the sound is coming from another source, then it may take some preparation to check the elements, and some post-scanning work to match up the elements. It's possible to do this in a manual way: marking the sync point on the image and sound film (if sound was on image element), matching it up visually before making renditions. Not all scanners can read soundtracks. In the digitization workflow, it may be necessary to incorporate .wav files coming from other sources that have been digitized (Dolby digital tracks, magnetic tapes, older sound formats). This is a "satellite workflow" that needs to be developed to get the sound files in line with the image files.

Plan enough time to test the workflow

From original film element to output element, hopefully there will be enough time to set everything up, make a few test runs of the entire life cycle, make corrections and repeat- and then start the actual workflow. This could take a few months, as a modest estimate. It is also important to test all types and qualities of film material, and the many settings of the scanner, to find out how to get the best scan from a variety of material, and to find out what material will scan the best- it is good to do this before preparing all the material for scanning, as it may change the decision about the best element to use for the desired result.

Plan a good workspace for checking

It's ideal to have a space where the output of the scanner can be reviewed, alongside the original element if necessary. Especially with sound syncing, it is helpful to be able to go from the digital workstation to the Steenbeck and run the film element. It's good to have a physical space that is big enough to keep the workflow organized, even in times when there is a backlog in a particular step- incoming film to be scanned or cleaned, film waiting while the digital assets are verified, and film waiting to be re-scanned.

Plan a suitable data backup strategy. Include workflow data too

A safe strategy is to store archival asset data (digitizations, renditions and technical metadata) in two discrete locations. But it's also good to remember to backup all the "everyday project metadata", not just the archival asset data. All the worksheets and project documents, equipment settings and passwords need to be backed up as well somewhere.

Moving big data around takes time, and needs to be accounted for in workflow planning

Backing up and moving huge amounts of data takes a lot of time. A feature film stored as uncompressed DPX is about 1.5 Tb of data. Writing and transferring these large data sets takes hours. Rendering takes time for even the best contemporary computers. It takes hours to restore a film from LTO tape, or copy to and from a hard drive. Even access renditions can be quite big- A mezzanine file might be about 150 GB. On the access side, there are questions about how to deliver it to a client/user- do you FTP it, send it over wire, or put it on a hard drive? Hopefully these questions will become less of an issue with technological improvements.

If possible, when something goes wrong, stop and fix the problem

Creating a backlog of half-done, or "problem scans" will only be more difficult to fix later on when the money and time run out.

Quality control and checking

How do you know if a scan is good or bad? Training staff to spot the common scanner errors is important. It is possible to run QC software on all the renditions. That takes time to set up the parameters, so that a computer can spot the errors. This is a nice option for a major scanning project with lots of similar assets. Remember that not only does the scan itself have to be checked, but also the sound, the renditions, all the tapes (or whatever storage media is used), and all the metadata, too.

It helps if the IT department is a willing participant in the scanning project

If the archive staff is not familiar with networking systems, hard drives, LTO players, film editing software or whatever is being purchased to do the project, then the IT department should be able to support them. Having a good relationship with the IT department, and having them available during installation of equipment will be necessary. Someone will need to be the "Administrator" for all the equipment, and be prepared to run required updates, do maintenance and have an oversight of passwords, versions and the networking.

Metadata and cataloging of the project

A scanning project will naturally result in a lot of digital data. Suddenly the archive is growing in a new direction. Does your current film catalog database support digital assets? If the digitization staff is not the same as the archive staff, is someone prepared to accession all these new assets, and does someone understand how to describe them, label them, organize them, use them and retrieve them?

Plan for retrieval of material

Probably the moment the digitization project starts, people will want access to the newly scanned material. Of course everyone is excited about the new digital files and can't wait to use them. If this is part of the project plan, then some staff needs to be able to retrieve digital material and respond to requests for material in contemporary formats (hard drives, DVDs, streaming files). As the amount of digital material grows, and word gets out about the possibilities associated with using a digital collection, this work will probably also grow as well.

Migration

Now that the files exist in the archive, they will need to be migrated in an appropriate number of years. For some forward-thinking planners, it might be possible to include this work into the digitization project budget. Remember to include the cost of the time it takes to organize, verify, re-label and re-verify the data, and update the catalog records-not just the time it takes to copy the data to the new media.

Presentation and documentation

Once the project is running, the digitization project team will probably be called upon to discuss and explain the process to other archives and visitors. During setup and subsequent activities, plan some time to document your progress (successes and failures, screenshots and sample scans) in a way that it can be explained to others, so that we can all learn from each other's experiences.

FIAF Technical Commission 2012