



"The world's largest online family history resource,"<sup>1</sup> is renowned for its ability to help others uncover their heritage by providing access to historical records such as census reports, photographs, marriage records, newspaper articles and more. In this synopsis of a six-part series posted on [Tech Roots Blog](#), Ancestry.com senior software development manager [Michael Murdock](#) breaks down the organization's Image Processing Pipeline (IPP) – i.e. how original genealogical records become online digital files. As featured in part four of this series, Ancestry.com utilizes [Mekel Technology](#) MACH-series [microfilm scanners](#) to digitize archived records. These Mekel scanners replaced almost twice as many competitive models while increasing output by 35%.

## PART 1 THE GOOD, THE BAD AND THE UGLY

This introductory post addresses how images arrive at the Ancestry.com Content Pipeline, which includes the Image Processing Pipeline that is the focus of the following posts. Images and materials collected by the firm's content acquisition team in paper, microfilm, digital and other forms arrive in various conditions, ranging from unblemished to barely legible.

To read the blog post in its entirety, [click here](#).

## PART 2 LIVING IN THE MESOSPHERE

Expanding on the subject of content, "Living in the Mesosphere" discusses the incredible volume of images that are processed at Ancestry.com. Murdock uses the analogy that if each image were a sheet of paper stacked on top of one another, that stack would be 31 miles high and reaching into the mesosphere. This content volume dictates the tools and technology used throughout the Image Processing Pipeline. The Mekel [MACH-series](#) is built specifically for this level of production-volume digitization.

To read the blog post in its entirety, [click here](#).

## PART 3 WHERE DO IMAGES COME FROM?

There are many challenges to image digitization that can result from "destructive operations" including: the quality and age of the paper, ink or film; improper handling of materials causing folds, tears or bows; and improper equipment settings such as camera exposure and compression. The Ancestry.com team follows a set of strategies to deal with these challenges and to make the processing pipeline simpler.

To read the blog post in its entirety, [click here](#).

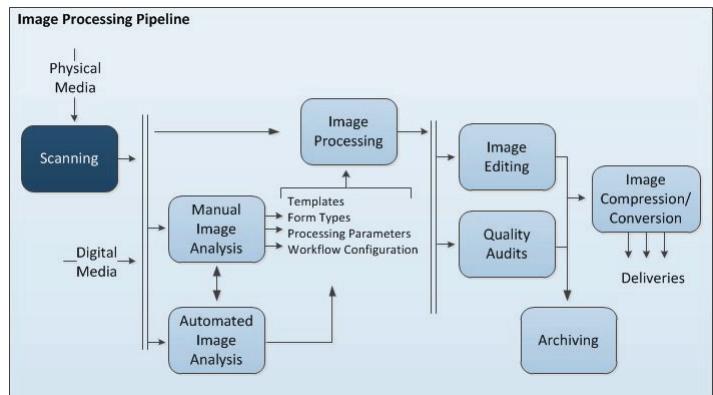
## PART 4 MICROFILM SCANNING

Part 4 of this series is reprinted here excluding several images and captions to conserve space. The original blog in its entirety can be viewed [here](#).

This post is the fourth in a [series](#) about the Ancestry.com Image Processing Pipeline (IPP). The IPP is the part of the content pipeline that is responsible

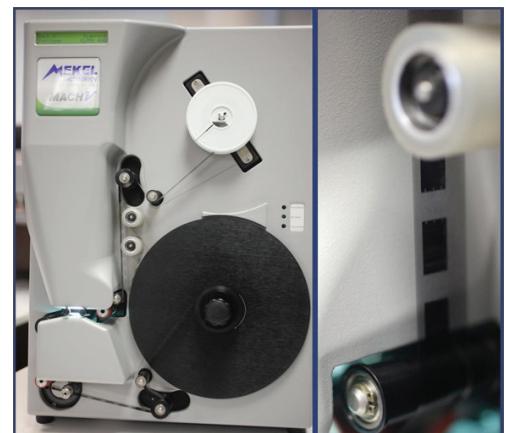
for digitizing and processing the millions of images we publish to our site. In this post I will present a bit of information about our microfilm scanning process.

A high-level depiction of the IPP is shown in the following diagram. Scanning, shown in the dark blue box, is the first step in the pipeline and is the process by which we convert media ([microfilm](#), [microfiche](#), paper) into digital images.



The Image Processing Pipeline – The Scanning Process is highlighted in the dark blue box.

The following photo panel shows a Mekel MACH5 microfilm scanner on the left and on the right a strip of the microfilm as it streams past the camera's CCD sensor. Although we more typically process 35mm film, in this photo we are scanning 16mm film.



Mekel MACH5 microfilm scanner

<sup>1</sup><http://corporate.ancestry.com/>

The following photo is a composite showing in the left panel a roll of 35mm microfilm. The film is shown in the right panel zoomed in to the first four frames on the film.



A roll of microfilm. The right panel shows a zoomed-in portion of the film.

scanner contains an 8,192 pixel CCD array that can scan between 80 and 160 megapixels per second. The internal pixel representation is a 12-bit grayscale depth, which allows for a tremendous amount of flexibility in adjusting the dynamic range for the conditions on the film.

The most interesting point here is that this process is creating fixed-sized image strips. In the past, the scanners we used would segment the frames from the film as it scanned. In other words, the scanner created the frames as it scanned and you were pretty much stuck with the segmentation it gave you. But with strip scanning the scanner produces fixed-sized strips and thus defers the segmentation to a subsequent framing step that is much more accurate in the way it identifies frames. More importantly, by deferring the segmentation we can involve a human reviewer who can be much more deliberate and thus more accurate in determining how the content on the film should be framed.

The relationship between strips and frames is shown in the following diagram. On the left of the diagram are the strips produced by the Mekel scanner. On the right of the diagram are the frames created from these strips.

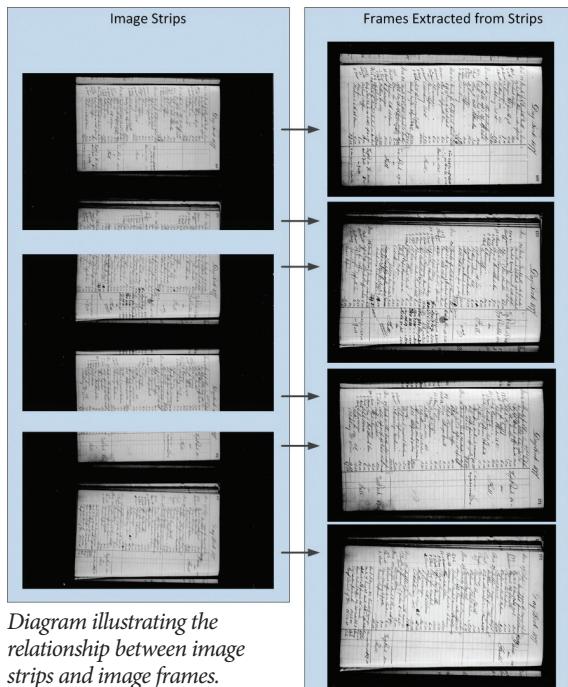


Diagram illustrating the relationship between image strips and image frames.

In this example, a roll of microfilm was scanned into 1367 strips, each 4096 pixels high. After an operator reviews and fine-tunes the scanner-supplied segmentation, 1837 image frames were extracted by stitching together the appropriate strips.

You have probably never even once wished you knew more about microfilm scanning technology. Creating 35mm rolls of microfilm is a nearly 80-year-old technology and microfilm scanners have been around for decades. But if you care (deeply) about producing high-quality images, getting this part of the process right is absolutely critical. Strip scanning is a fairly recent development, and the work we have done the last few years to do the stitching of strips into frames on our server farm has been something of a minor break-through, enabling the IPP to produce both higher volume and higher-quality images.

## PART 5 AUTO-NORMALIZATION

This post covers the processing operations that take place after the images are digitized. Source images enter the Image Processor to be checked for pixel distribution, auto-normalization and image contrast. The images then continue on to the Image Quality Editor.

To read the blog post in its entirety, [click here](#).

## PART 6 AUTO-SHARPENING

The final post in this series on Ancestry.com's IPP is a continuation of part five's discussion on the core image processing operations, particularly auto-sharpening. This operation attempts to enhance the image by amplifying the high-frequency components of the image, such as text edges, to remove any blurring effects.

To read the blog post in its entirety, [click here](#).

**Crowley note:** As mentioned in these blog posts, the Mekel Technology [MACH5](#) is a production-level microfilm scanner capable of scanning at speeds of up to six minutes per roll with a range of output options including TIFF, JPEG, JPEG2000, PDF, PDF/A and more. Released in 2013, the Mekel [MACH12](#) features a 12-bit camera with full 12,288 pixel CCD array for true optical resolution of up to 750 dpi. All MACH-series microfilm scanners are powered by proprietary [Quantum](#) software, offering a range of image enhancement tools and simple processing workflow.

## ABOUT ANCESTRY.COM

Founded in 1983 as a publishing company, [Ancestry.com](#) is dedicated to helping people discover their roots by providing access to a vast repository of family history records dating back to the late 1300s. Since their website launch in 1996, Ancestry.com has digitized and hosted over 175 million photographs, documents and written stories, garnering approximately 2.7 million subscribers across their range of family history sites.<sup>2</sup>

## ABOUT THE CROWLEY COMPANY

The Crowley Company is a world-leader in digital and analog film technologies and provides an extensive number of digital document and film conversion [hardware](#) and [services](#) to the academic, publishing, commercial, government and archive sectors.

<sup>2</sup><http://corporate.ancestry.com/about-ancestry/>