The Family and Church History Department Imaging Process and How the Digital Camera System Design Fulfill Requirements Current Status and Needed Development

Rick Laxman

Note: Areas needing further development are indicated with an asterisk (*)

Negotiations and Project Administration

- ➔ Permissions
  - Rights to duplicate and distribute
    - Intranet
    - Internet
    - Paper
    - Any other technology or method developed in the future
- ➔ Derived work issues—need a legal definition of what constitutes a “derived work” when creating, enhancing, compressing, or indexing a digital image of a document*
- ➔ Provenance
  - Document appearance at time of capture
  - Record location and order/sequence
  - Enhancements
- ➔ Win/Win/Win for all parties involved
  - Archives, Patrons, Genealogical Society of Utah
- ➔ Provide better access to images/information
- ➔ Managing Imaging Projects*
  - Need the hardware, software tools, and databases to effectively manage worldwide imaging projects*

Capture, Convert and Acquire

- ➔ Equipment used for GSU digital capture
  - Digital Camera—the GSU utilizes two digital cameras. The prototype system includes the Roper Megaplus 6.3i camera, that uses a 6.3 megapixel, CCD sensor array. The second system that will be used by the GSU has, as it’s centerpiece, the Atmel, Camelia 8 camera, which uses an 8 megapixel, CCD array.
  - Area, or matrix arrays are much faster due to the fact that the entire image can be captured at once—much like a camera that captures a scene on a piece of film.
  - Conversely, scanners typically use linear arrays that capture information one line at a time.
  - Scanners yield much higher resolution.

- The computer system used in the imaging system is optimized for production speed.
  - 384-MB RAM
  - The present systems utilize a one GHz processor.
  - The two 36-MB hard drives have been striped: RAID 0 Configuration
  - The above configuration will yield the following:
    - 24-bit Color Image~6 to 8 seconds
    - 8-bit Grayscale Image~1 to 2 seconds
    - 1-bit Bitonal Image~2 to 3 seconds

➔ Some considerations for capture:

- Spatial Resolution is normally measured in DPI. Due to the nature of the GSU Imaging system, dpi becomes a relative characteristic. DPI is dependent on the focal length of the lens, and the distance of the camera's CCD array from the document being captured.
- The GSU has defined resolution as "the ability to capture the thinnest line segment with 3 to 4 pixels per line segment (ppls), with a signal to noise ratio of no less than 1:1.5; for an 8-bit grayscale and/or 24-bit color image.
- These measurements occur during the initial certification of the imaging systems.
- Depth Resolution
  - The GSU, digital camera has the capability to image in three bit-depths:
    - 1-bit bitonal—black and white
    - 8-bit grayscale—256 shades of gray
    - 24-bit color—over 16 millions colors
  - The bit depth determines the resultant file size:
    - 1-bit bitonal = 150k to 200k
    - 8-bit grayscale = ~6.0 MB
    - 24-bit Color = ~18.0 MB
  - Image Compression
    - Image compression is based on the ultimate use of the images. The GSU uses industry standard image formats.
      - ♦ JPEG: Intra- and Internet use.
      - ♦ Tiff: "master" images
  - Image Control
    - The imaging software used by the GSU allows absolute exposure and color-balance control over the resultant images. By proper exposure color-balance, the operator can optimize the amount of information captured by the digital camera.
    - Presently, there are no enhancements that occur simultaneously with image capture. Any enhancements that can be made to images are post-capture. Future versions of GSU imaging software will include simultaneous enhancement capability; i.e. image deskew, etc.

- Image Quality
  - GSU imaging software allows the operator instant feedback on image quality. If images are substandard at the time of capture, the operator can adjust the camera settings, and perform a retake of the image in question.
  - Readability of information on a given document/image by a user/patron is the ultimate requirement for all of the images captured by the digital imaging system. This requirement is measurable using the following criteria for a draft of the proposed Family and Church History, Digital Imaging Standard, version 1.0:
    - ♦ If a digital, technical target is available, the capture device used for digitizing the document shall be capable of representing a 0.1 mm line segment with a reflective density of 0.3 as 3 to 4 pixels in the grayscale domain and 2 to 3 pixels in the bitonal domain.
    - ♦ If a digital, technical target is not available, the thinnest, lowest contrast line segment of a character in a document series shall have a minimum of 3 to 4 pixels in the color or grayscale domain, and 2 to 3 pixels in the bitonal domain.
    - ♦ Agreement shall be the third test of resolution. "Agreement" simply stated is that all line strokes or characters that are visible on the document being captured shall be legible on the digital image.

➔ Quality Audit*

- The GSU will perform a statistically random audit on all incoming images using software created for the GSU.
- Audit software application is used to select, view, and record image quality of a random sample of images
  - A folder of image, equivalent to a single volume or one roll of microfilm, is selected for an evaluation audit
  - A quality audit is done on multiple folders to verify the work of the evaluation auditors.
  - Reject codes are selected and attached to any image that does not meet the quality standard
  - Since image quality evaluation takes place at the time of image capture (original documents or scanned microfilm), a low threshold (currently 3 rejected images per folder) is selected,
  - Once a folder is rejected, a retake request is sent to the digital camera or scanner operator to re-image the documents represented in the folder

Image and Metadata Processing

➔ Metadata: five categories

- Administrative
- Discovery or Description

- Preservation
- Technical
- Use

- Technical metadata is captured simultaneously with images
  - Operator Name
  - Date and Time of Capture
  - Exposure information
    - Gain
    - Black Level
    - Shutter Speed/Exposure Time
  - Descriptive information about the record being captured.
    - Archive
    - Locality of record
    - Dates covered in record
    - Etc.
- Captured metadata becomes operator report*

Storage & Preservation

➔ Metadata stored in database linked to images
➔ Images stored in image file server in a file/folder structure
➔ Storage and Migration*
- Providing Quality in Storage and Preservation
  - Storage Environment
    - Climate controls and security
  - Backup
  - Hardware/Software Obsolescence and Migration needs to address and develop approaches to the follow elements:*
    - Technology Migration*
    - Backward Compatibility*
    - Migration window of opportunity*
    - Image, metadata, and linking comparison on old and new media, software, & hardware*
  - Media Shelf Life
  - Media Inspection*
    - A system needs to be developed to test and verify media to determine the need to copy images and data to new media
  - Media Migration
  - People Skills Migration and training

➔ Preservation*
- Preservation of electronic format
  - Images
  - metadata
  - Database
  - links between images and index data

- Preservation in human-viewable format
  - Microfilm is a viable, but not cost effective option today*
  - Other technology/media needed to be developed to meet this essential requirement*
  - Both images and index data need to be preserved on a human-viewable technology/media*
    - The images and indexes that point to those images should be preserved jointly on the same piece of human-viewable media.
    - The indexes should be written in a font and format to allow for the recreation of the indexes in a database through scanning and OCR technology
    - The images written on the human-viewable media should include the image file name (for OCR) to re-create the links between the indexes and images after the scanning of the images and the scanning and OCR processing of the human-viewable indexes is completed

Indexing and Cataloging

➔ Indexing to the name level

- Labor-intensive process*
- Look at using the Internet*
- Need automation/technology to speed up the process*
- Three scenarios for indexing images:
  - Expand the volunteer labor pool by orders of magnitude to do manual indexing of images using the Internet
  - Develop software tools to automatically recognize and categorize handwriting in various languages to create indexes
  - Use a combination of an expanded volunteer labor pool and software tools to exponentially increase the rate of indexing

➔ Catalog to the volume level

- Describe locality, time period, record source, event type, and other general information
- Can be linked to folder containing images for the specified volume
  - could be presented as “digital microfilm” without a name index

Access and Distribution

➔ Access is provided through three tools—separately or combined*

- Research Guidance—step-by-step guidance on searching for your ancestors
- Catalog—The Family History Library Catalog is the key to searching the world's largest collection of genealogical records, including the names of more than 2 billion deceased people collection.

- Index to Images—Index to names associated with events, locations, and dates

➔ Distribution to be in electronic format and paper*

- Electronic format via the Internet to
  - Family History Library*
  - Family History Centers*
  - Homes*
- Access and security issues at each level need to be addressed*
  - Compliance with contract signed with original record/copyright holder
    - The ability for a user to view, print, or download an image for personal use*
    - These restriction could be based on user location or user status (member of an organization) *
- Possible scenarios for paper distribution *
  - Send images to Family History Centers, Wards, or Stakes
    - Images printed by local units and distributed to members and patrons
  - Print images at Church headquarters or in Area Offices and distribute to members and patrons