

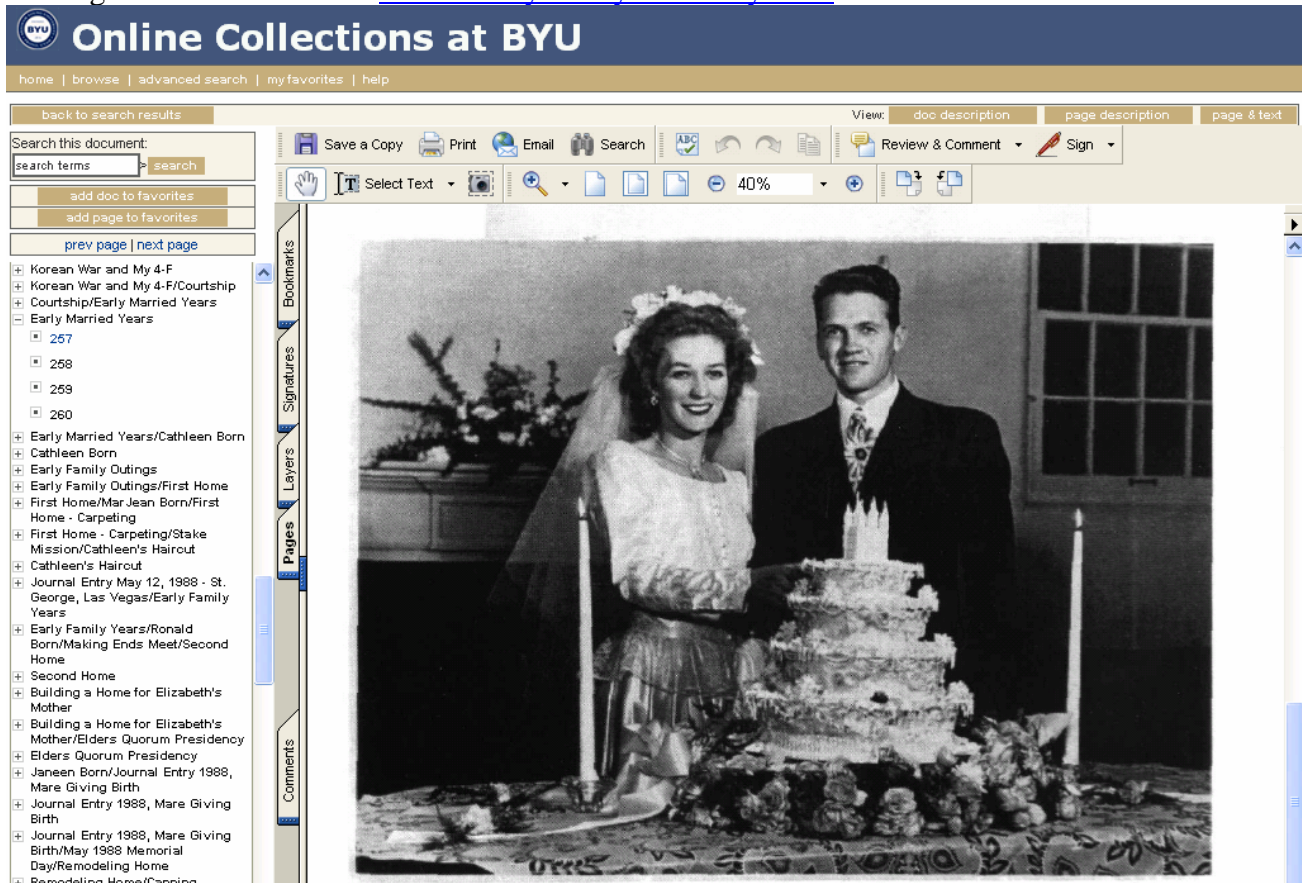
Moving Production In House

Jeri Jump
Emerging Technology
Family History Department, LDS Church

1. Problems and Solutions

This presentation outlines the problems and solutions for bringing the digital production of the Family History Archive collection in house.

❖ Image from the collection www.familyhistoryarchive.byu.edu

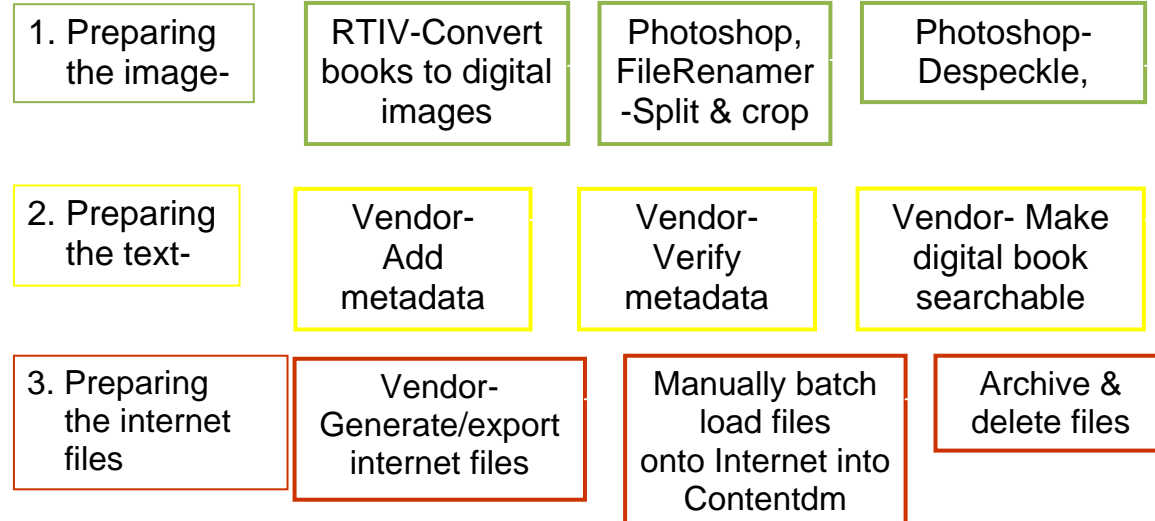


The screenshot displays the 'Online Collections at BYU' website interface. The top navigation bar includes links for 'home', 'browse', 'advanced search', 'my favorites', and 'help'. Below this, there are options to 'back to search results', 'Save a Copy', 'Print', 'Email', 'Search', 'Review & Comment', and 'Sign'. A search box is present with the text 'Search this document: search terms'. On the left side, there is a sidebar with a tree view of document categories, including 'Korean War and My 4-F', 'Courtship/Early Married Years', 'Early Married Years', 'Cathleen Born', and 'Journal Entry 1988, Mare Giving Birth'. The main content area shows a black and white photograph of a bride and groom standing behind a large, multi-tiered wedding cake. The bride is wearing a white dress and veil, and the groom is in a dark suit. The cake is decorated with flowers and has several lit candles. The website interface also includes a 'Pages' sidebar with 'Signatures', 'Layers', and 'Comments' sections.

2. What We Learned from the Pilot Project

- Using one tool could reduce the time, money, and the hassle it took to manage and license multiple software applications and vendors
- It was expensive and time consuming to use a vendor
- We wanted more of the process to be automated
- We could streamline the process

3. Phases of Pilot Production Process with Software



4. Tool Problem

Because many of the tools belonged to the vendor, we needed to find or produce our own tools.

- a) Page processing and batch processing
- b) Speed up production
- c) Provide a way for a non-technical person to do a technical processes
- d) One tool

5. Tool Solution

- a) The Church had already purchased the Kofax Ascent Capture software for some of its other digitizing needs.
- b) We began testing and learning the software to see if it might be a solution for our project as well.
- c) The software proved to be robust out of the box, and customizable by nature. It became our solution.

6. Software Problems

1. Scanning- Scans coming from multiple sources such as scanners, overhead cameras, and microfilm. We wanted to be able to import these images and also to scan directly into the software. The images being imported are 200-400 dpi grayscale tif images.
2. Cropping/Splitting files- Crop off excess space around the edge of the page and also split 2 images in one file. We also wanted to be able to capture documents that spanned multiple pages and split multiple documents occurring on a single page.
3. Cleaning-If an image is poor quality it may need the contrast adjusted or need some de-speckling done to obtain acceptable OCR output.
4. Adding Metadata- Add metadata for each page.
5. Validating numbering- Validate that no pages are skipped, that they are in the correct order, and that each page has a number assigned to it.

6. Creating text searchable files- Output text searchable PDF files in both a single & multi-page format for Contentdm import.
7. Creating Contentdm ready files- Package PDF files and a tab delimited text file with the 40+ fields of necessary information for building a compound object in Contentdm.
8. Portability- Need software that is portable so that it can be setup and administered at a central location, and distributed to remote sites world wide.

7. Software Solutions

1. Standard Kofax Ascent Capture Software
 - a. Scanning, adding metadata, creating text searchable files, portability-
2. Customized Modules
 - a. Book Cropper - Cropping/Splitting files
 - b. Image Processing - Cleaning files
 - c. Page Numbering Validation- Validating numbering
 - d. Custom Release Script - Creating Contentdm ready files

8. Book Navigation Problems

1. Each hard copy of a book contains information that helps the reader navigate through the book.
 - a. For example: Table of Contents, Chapter headings, Index headings, the position information of the pages themselves and the order they come in
2. The reader needs this navigation information or it's difficult to use the book or find specific items in the book.
3. There is an even greater need for navigation information in a digital book. Since the reader only has the digital files, they lose the position information provided by the hard copy itself.

9. Book Navigation Solutions

The Digital Table of Contents

1. This is the most time intensive part of the process.
2. It has to be done manually
3. Each page has to be looked at
4. It can be complex depending on the layout of the book
5. Sometimes-
 - multiple "books" are combined together to make one book
 - odds and ends are added together to make a "book"
 - no page numbers are included
 - some page numbers are included but others are missing
 - the table of contents from the author doesn't match the layout of the book
 - there is no table of contents
 - we have a film copy that has pages missing, and we don't have the original

10. Digital Table of Contents Pagination Worksheet

❖ Digital Table of Contents Worksheet

	A	B	C
1	Page Number	Page Heading	Chapter Heading
2	0	Title Page	Front Matter
3	i	Acknowledgements	Front Matter
4	ii		Front Matter
5	iii	Table of Contents	Front Matter
6	iv		Front Matter
7	1	Gold/Parker Pedigree Sheet	Front Matter
8	2		Front Matter
9	3	Introduction	Front Matter
10	4	Some Utterings	1 Milton Osborne Peck Gold
11	5		1 Milton Osborne Peck Gold
12	6		1 Milton Osborne Peck Gold
13	7		1 Milton Osborne Peck Gold
14	8		1 Milton Osborne Peck Gold
15	9	Autobiography	1 Milton Osborne Peck Gold
16	10		1 Milton Osborne Peck Gold
17	11		1 Milton Osborne Peck Gold
18	12		1 Milton Osborne Peck Gold

11. The Digital Table of Contents Online

❖ Digital Table of Contents from Online Book

previous page : next page

The descendants of John Perry of Roxbury, MA : listing the first five generations with the children comprising the sixth generation including the full descent of the compiler's sixth generation ancestor, Abner Perry of Holden, MA

Page number

Chapter heading

Page heading

- Front Matter
- ▣ i- Title Page
- ▣ ii- Donation Page
- ▣ iii- Author Letter
- ▣ iv- Author Letter
- ▣ v- Map
- + Generation One
- + Generation Two
- + Generation Three
- + Generation Four
- + Generation Five
- + The Descent of Abner Perry #91
- + Printing Version 6 MB

12. Implementation Problems

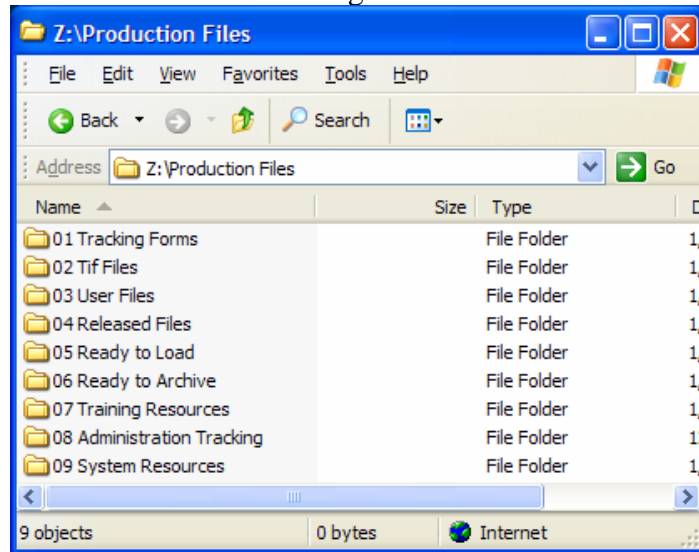
We learned from the pilot project and looked for areas to make improvements.

1. Many steps in the production process
2. Complicated file structure
3. Complicated tracking system

13. Implementation Solution

We reduced the steps, simplified the folder structure and tracking system.

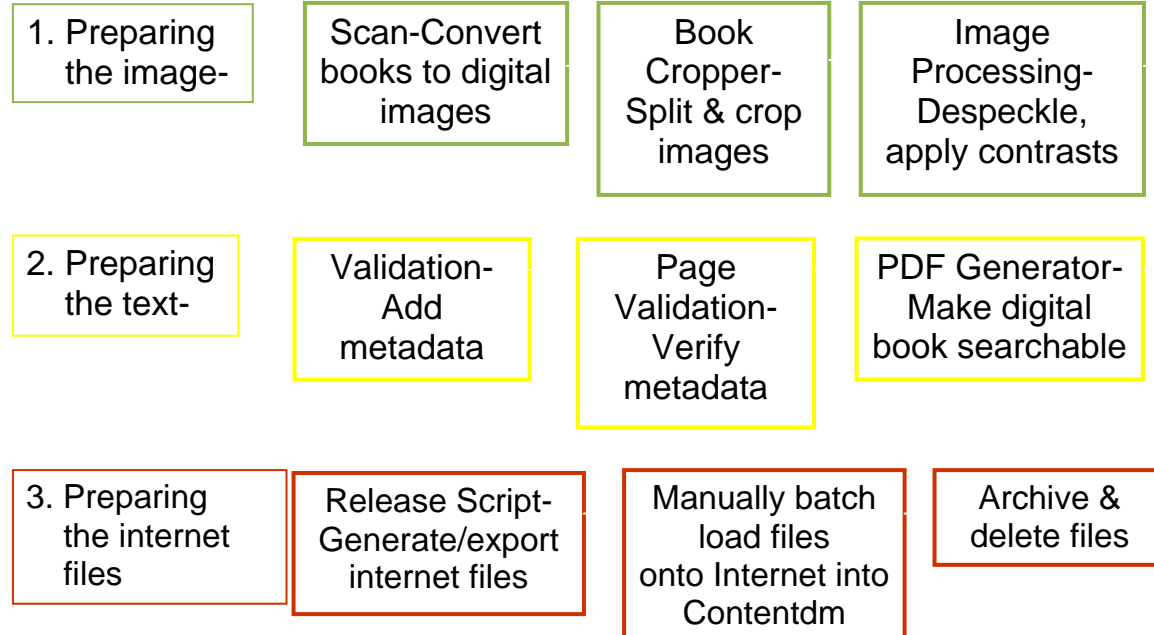
❖ Screenshot of Chronological Folder Structure



❖ Tracking Form Page

	C	D	I	J	K	L	M
	Title	Author	File Name	Scanned by	Machine #	Scan Date	Pages
	The Descendants of the Finks Family	Finks, Betty Lou	TN-778889	Tolman	1-SLC	11/29/2006	111
	Dertien Family	Mosser, Ruby	TN-703353	Tolman	1-SLC	12/13/2006	440
	Netherton's Beginning	Helmer, Dorothy	TN-1351892	Tolman	1-SLC	12/5/2006	308
	From Tingewick to Tioga	Holton, Joseph T.	TN-1326135	Tolman	1-SLC	11/29/2006	198
	Our Eccles Family	Thomas, Myrella	TN-780343	Tolman	1-SLC	11/30/2006	145
	The Descendants of Thomas and Ruth Faith of Somerset County, Pennsylvania	Hays, Samuel P.	TN-729241	Tolman	1-SLC	12/1/2006	268

14. Phases of In House Production Process with Modules



15. Progress Report

1. Custom modules are being created now
2. We can produce books now, but part of the process has to be done manually. We won't go into full production until we get the custom modules.
3. We plan to distribute the software and production process to remote libraries so they can digitize their own collections using volunteers starting in 2007.
4. Our goal is to produce 10K book the first year, 20 books each year after that
5. We estimated about 2 man hours to produce an average book.
6. We are staffed by both full time missionaries and Church service missionaries
7. We are set up in the Church's Orem office. Our facility is called the Orem Digital Processing Center

16. Contact Information

- ❖ Jeri Jump
- ❖ jumpjl@ldschurch.org
- ❖ 801-787-7413 cell
- ❖ 801-240-7673 work

- URL to Family Search: <http://www.familysearch.org/>
- URL to Family History Archive: www.familyhistoryarchive.byu.edu