



The Mormon Diaries Project

Scott Eldredge, Digital Initiatives Program
Manager

Harold B. Lee Library

Frederick Zarndt, CTO
iArchives

What Is Transcription?



- ◆ **Transcribe** v.t. 1. To write over again; copy from an original. 2. To translate into standard written form.
- ◆ **Transcription** n. 1. The process or act of transcribing. 2. Something transcribed.
- ◆ **Transcript** n. 1 Something transcribed.

Character Recognition



- ◆ Optical Character Recognition (OCR)
 - Machine-print, block characters only
 - Results depend on image quality
- ◆ Intelligent Character Recognition (ICR)
 - OCR for handprint or handwriting
 - Online: Characters detected when written
 - Offline: Characters detected after written
 - Rejean Plamondon and Sargur N. Srihari, “*On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey*”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 1, January 2000

Unconstrained Handwriting

John Stillman Woodbury



Recorded
120 baptisms by John S. Woodbury
from July 24, 1854 to July 24, 1858
in Cadmus conference, while under his
Superintendency.
1 marriage
29 children blessed -

Thursday 10 Road rough over the moun-
tain, when we got down into Little Salt
Lake valley we had good road. About
noon came to Elk Horn Spring. The cattle
wer very thirsty but we could scarcely
get a taste for them all the Spring being
so small. While here Elder George A.

Transcription of Handwriting



- ◆ Poor results from algorithmic transcription of unconstrained handwriting
- ◆ Manual transcription
- ◆ Few, but diverse transcription projects
- ◆ Internet distribution and collection of digital images and transcribed text
- ◆ Establishment and management of transcription workflow process is significant barrier

Project Gutenberg



- ◆ Oldest producer of free electronic books on the Internet
- ◆ Volunteers produced 15,000+ eBooks
- ◆ OCR correction from digital text images
- ◆ Mostly plain text but also HTML, PDF, TeX, Postscript
- ◆ <http://www.gutenberg.org/>
- ◆ Volunteers sign up and download images and upload transcribed text at <http://www.pgdp.net/c/default.php>

Early English Books Online Text Creation Partnership



- ◆ Partnership of University of Michigan, University of Oxford, Council on Library and Information Resources (CLIR), ProQuest Information and Learning, and others
- ◆ Structured SGML/XML text editions for a portion of the Short Title Catalog of Early English books published between 1473 and 1700
- ◆ Target transcription accuracy of 99.995%
- ◆ Transcribed text validated against DTD
- ◆ Transcribed text linked to digital images
- ◆ <http://www.lib.umich.edu/tcp/eebo/>
- ◆ <http://eebo.chadwyck.com/home>

Project Runeberg



- ◆ Project of [Linköping University](#) in Sweden
- ◆ Internet's biggest center for Nordic literature
- ◆ Raw OCR text presented with digital image
- ◆ Readers may submit corrections to OCR text online
- ◆ Moderator accepts/rejects corrections
- ◆ <http://runeberg.org/>

American Pioneer Diaries 1



- ◆ University of Utah, Utah State University, Utah State Historical Society, and Lee Library transcribed 49 handwritten pioneer diaries (Library of Congress grant)
- ◆ Approximately 30,000 pages from 49 diaries transcribed and XML tagged to TEI schema with Wordperfect and XML Spy
- ◆ <http://overlandtrails.lib.byu.edu/>



Wednesday, January 1st, 1862. 3

Emma and I started this morning a little after 7 a. m. for Birmingham. When we reached the landing stage, we discovered that the fog was so thick that only one boat was allowed across the ferry at one time, while, in ordinary cases, about 5 generally run from each side of the river. Bells were rung each [page torn] the whistle of the boat was kept up occasionally [page torn] The object in ringing the bells was to guide the steerer [page torn] he would not know where to steer [page torn] it not been for [page torn] such contrivance, for it was impossible to see more than a few yards before. The object of [page torn] to give a signal to the vessels that were lying on the river, that they might ring their bells in time to give a [page torn] signal to the boat to avoid a collision. This of course, caused the boat to be very much detained, [page torn] the consequence was [text runs off the page] that we did not arrive at the Station until about 1/2 past 8, while the train started at 10 minutes past 8. We talked the matter to the station-master, who granted us a ticket to go with the next train (2d. Class, 10 minutes to 10,) at third-class fare. I had an interesting talk with [page torn] gentleman on the train respecting the American War, and, finally, respecting "Mormonism". He was very fair; but, although he respected me, he would not receive our principles. I arrived at bro. John. Sabin Smith's, 12 1/2, St. Paul Square, Birmingham, about 3 p. m. The General Council of the Authorities of the European Mission commenced at [page torn] o'clock this morning, at Farm St. Chapel, Hockley, Birmingham, but did not {save} have the privilege of meeting with them ^

Wednesday, January 1st, 1862. Emma and I started this morning a little after 7 a. m., for Birmingham. When we reached the Land- ing - stage, we discovered that the fog was so thick that only one boat was allo [page torn] ross the ferry at one time, while, in ordinary cases, about 5 generally run from each side of the river. Bells were rung each [page torn] the whistle of the boat was kept up occasionally [page torn] The object in ringing the bells was to guide the steerer [page torn] he would not know where to steer [page torn] t not been for [-]uch contrivance, for it was impossible to see more than a few yards before. The object of [page torn] to give a signal to the vessels that were lying on the river, that they might ring their bells in time to give a [page torn] cient warning to the boat to avoid a collision. This of course, caused the boat to be very much detained, [page torn] the consequence wa [text runs off the page] that we did not arrive at the Station until about 1/2 past 8, while the train started at 10 minutes past 8. We talked the matter to the station-master, who granted us a ticket to go with the next train (2d. Class, 10 ^ minutes to 10,) at third-class fare. I had an interesting talk wit [page torn] gentleman on the train respecting the American War, and, finally, respecting "Mormonism". He was very fair; but, although he respected me, he would not receive our principles. I arrived at bro. John. Sabin Smith's, 12 1/2, St. Paul Square, Birmingham, about 3 p. m. The General Council of the Authorities of the European Mission commenced at [page torn] o'clock this morning, at Farm St. Chapel, Hockley, Birmingham, but did not {save} have the privilege of meeting with them ^

The dates in **bold type** have been added for clarity .
Green text indicates insertion in original diary.
Red text indicates deletion in original diary.
Extraneous indicates material that appears to be extraneous to the dated entry.

Journals, 1861-1863
Ajax, William, 1832-1899
MSS 1488 [\[1\]](#) [\[2\]](#)

[page 1]

William Ajax
Journal [\[3\]](#)

[page 3]

1862

January 1862

January 1, 1862

[\[4\]](#) Wednesday, January 1st, 1862.

Emma and I started this morning a little after 7. a. m., for Birmingham. When we reached the Landing - stage, we discovered that the fog was so thick that only one boat was allo [\[page torn\]](#) ross the ferry at one time, while, in ordinary cases, about 5 generally run from each side of the river. Bells were rung each [\[page torn\]](#) the whistle of the boat was kept up occasionally [\[page torn\]](#) The object in ringing the bells was to guide the steerer [\[page torn\]](#) he would not know where to steer [\[page torn\]](#) not been for [-]uch contrivance. for it was impossible to see more than a few yards before. The object of [\[page torn\]](#) to give a signal to the vessels that were lying on the river, that they might ring their bells in time to give a [\[page torn\]](#) cient warning to the boat to avoid a collision. This of course, caused the boat to be very much detained, [\[page torn\]](#) the consequence wa [\[text runs off the page\]](#)

that we did not arrive at the Station until about 1/2 past 8, while the train started at 10 minutes past 8. We talked the matter to the station-master, who granted us a ticket to go with ~~the~~ next train (2d. Class, 10 minutes to 10,) at third-class fare. I had an interesting talk wit [\[page torn\]](#) gentleman on the train respecting the American War, and, finally, respecting "Mormonism". He was very fair; but, although he respected me, he would not receive our principles. I arrived at bro. John. Sabin Smith's, 12 1/2, St. Paul Square, Birmingham, about 3. p. m. The General Council of the Authorities of the European Mission commenced at [\[page torn\]](#) o'clock this morning, at Farm St. Chapel, Hockley, Birmingham, but did not ~~even~~ have the privilege of meeting ~~aw~~with them [\[5\]](#)

Overland Trails Text PDF



American Pioneer Diaries 2



- ◆ Workflow process and management not automated
- ◆ Labor costs high
- ◆ Work done at different locations
- ◆ Name normalization difficult
- ◆ XML tagging not standardized

Mormon Diaries 1



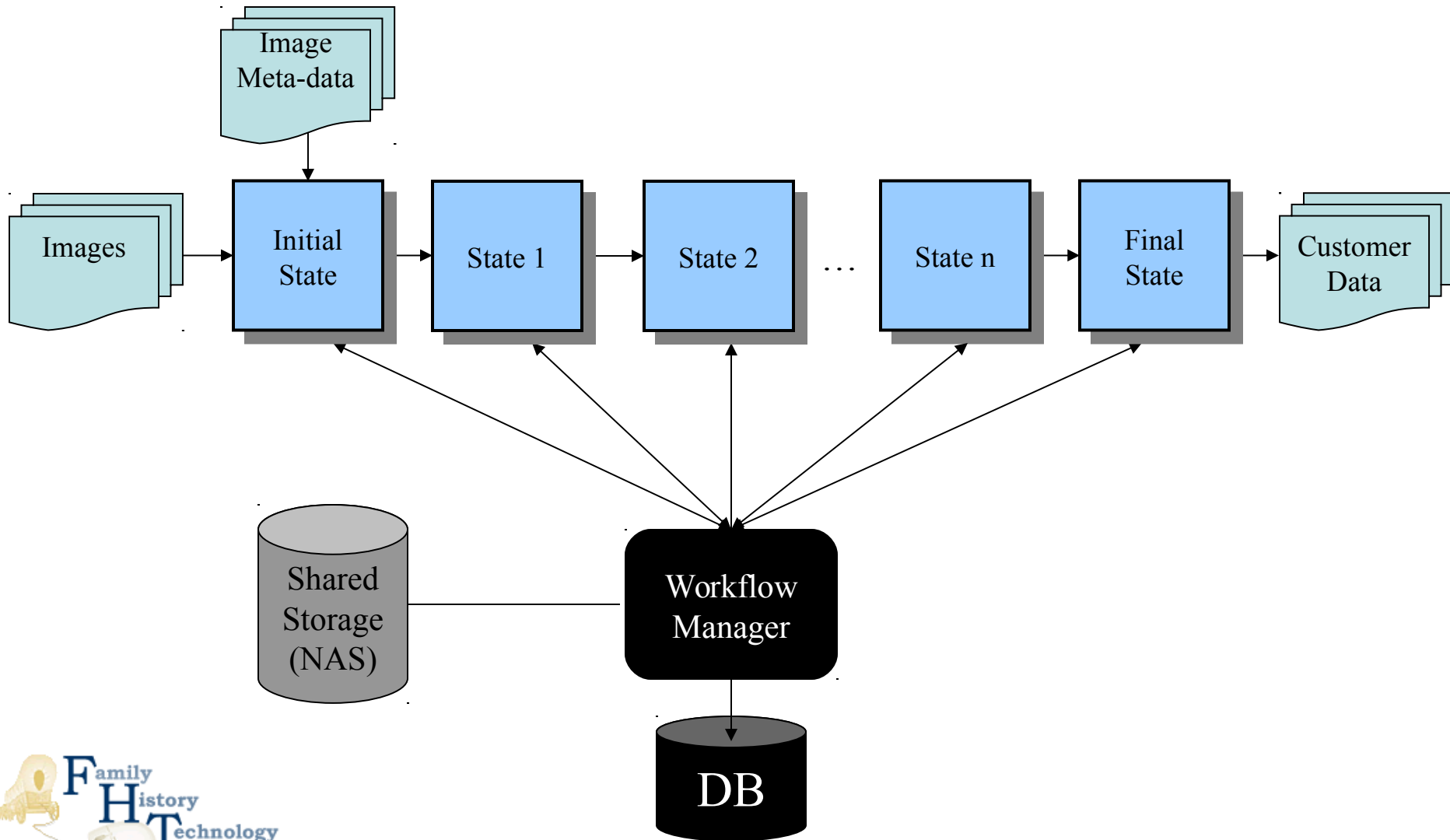
- ◆ Over a century of first-hand church history
- ◆ Scope of Mormon diaries project
 - 70,000 pages
 - 390 volumes
 - 116 diarists
 - 20 countries, 5 continents
- ◆ Scope of American pioneer diaries
 - 30,00 pages
 - 49 diarists

Mormon Diaries 2

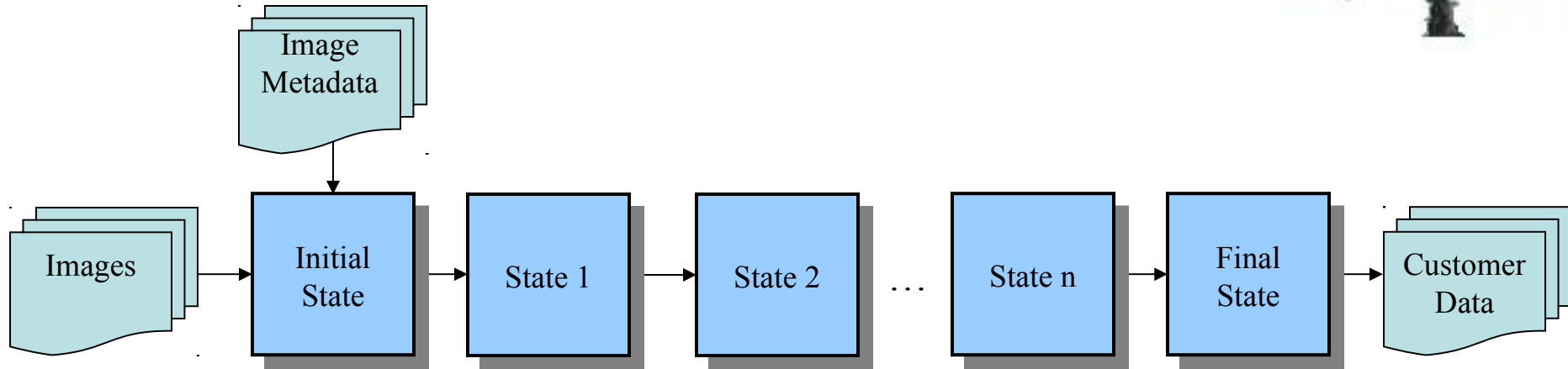


- ◆ Improve, automate, and streamline workflow
- ◆ Design software application for transcribing and tagging handwritten text
- ◆ Normalize work done at different locations and by different people
- ◆ Simplify name normalization and authority
- ◆ Transform transcriptions into diverse formats including TEI and PDF

State-based Workflow

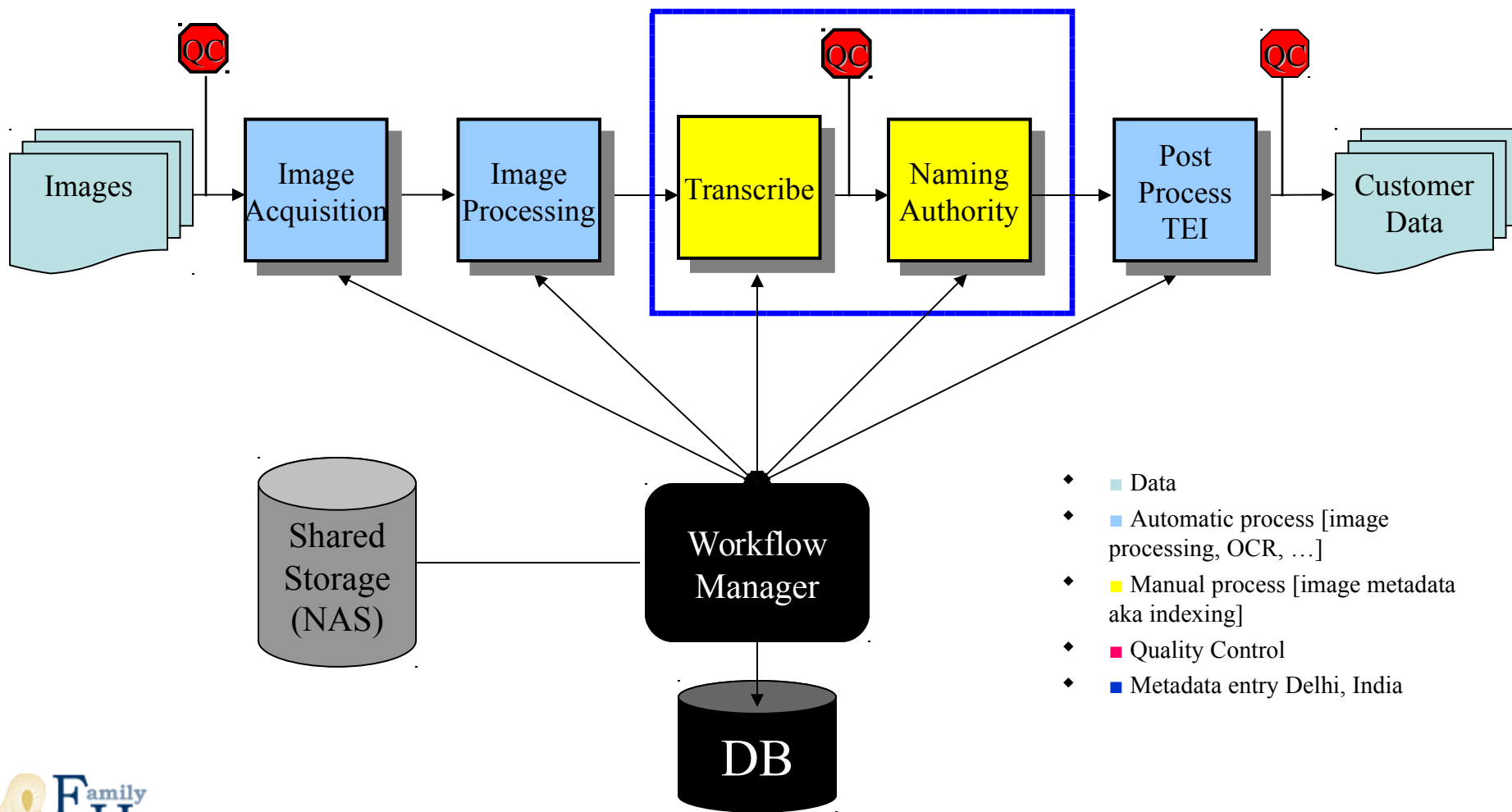


State-based Workflow

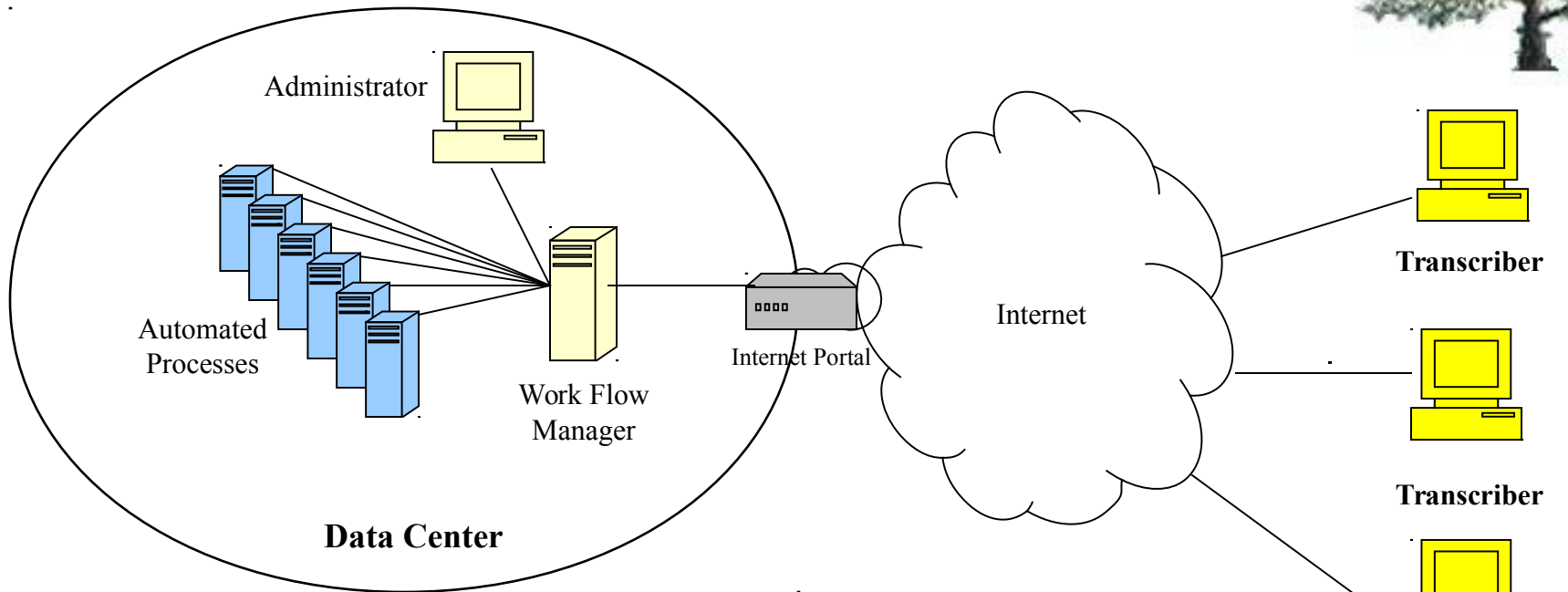


- ◆ State transitions are governed by the nature of the workflow
- ◆ Number and type of states is flexible and customized to the workflow
- ◆ States may be required or optional depending on workflow properties
- ◆ Each state has a driver specific to the workflow
- ◆ States may be blocking or non-blocking (dependent on the workflow and nature of the state)
- ◆ Quality control gates may optionally be configured to follow one or more states

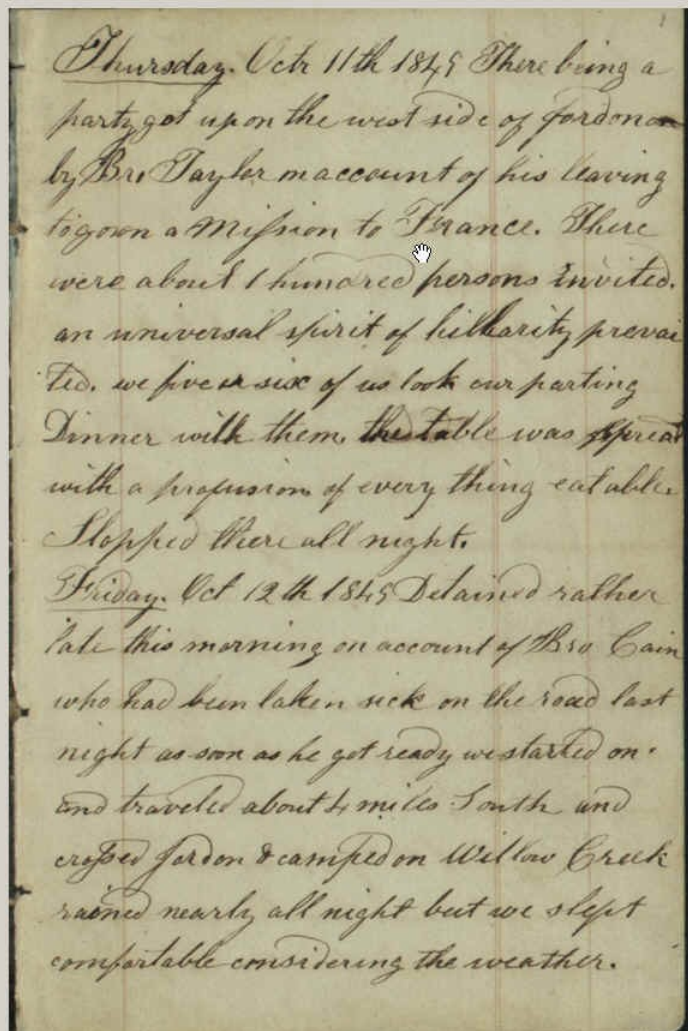
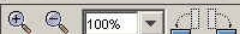
Mormon Diaries Workflow



Distributed Processing



- ◆ Work is distributed to computers hosting automated and manual processes by work flow manager
- ◆ Work scheduler is modular and can be easily changed as required
- ◆ Computers hosting automated and manual processes can do work after completing registration with the work flow manager
- ◆ Third party licensed software (if any) is hosted in data center: no license management problems.



1
 Thursday, Oct 11th 1849 There being a party got upon the west side of Jordan on by Br. Taylor on account of his leaving to go on a mission to France. There were about 1 hundred persons invited an universal spirit of hi charity prevailed. we five or six of us took our parting Dinner with them. the table was spread with a profusion of every thing eatable Stopped there all night.

Friday, Oct 12th 1849 Detained rather late this morning on account of Bro Cain who had been taken sick on the road last night us soon as he got ready we started on and traveled about 4 miles south and crossed Jordan and camped on Willow Creek rained nearly all night but we slept comfortable considering the weather.

Metadata Pages

000a
 a001_mss_1521_v01
 a002_mss_1521_v01
 b001_mss_1521_v01
 b002_mss_1521_v01
 b003_mss_1521_v01
 p001_mss_1521_v01
 p002_mss_1521_v01
 p003_mss_1521_v01
 p004_mss_1521_v01
 p005_mss_1521_v01
 p006_mss_1521_v01
 p007_mss_1521_v01
 p008_mss_1521_v01
 p009_mss_1521_v01
 p010_mss_1521_v01
 p011_mss_1521_v01
 p012_mss_1521_v01
 p013_mss_1521_v01
 p014_mss_1521_v01
 p015_mss_1521_v01
 p016_mss_1521_v01
 p017_mss_1521_v01

Summary



- ◆ Configurable workflow management system for transcription (and other) projects
- ◆ Configurable transcription application
- ◆ Flexible data tags and name normalization
- ◆ Painful stuff – workflow management – can be configured once and re-used



Questions?

