Management and Distribution of Historical Document Images

Stephen W. Liddle, Rayman D. Meservy
School of Accountancy and Information Systems

John W. Welch
BYU Studies
Brigham Young University
Charles Rich Collection

- 6 CD’s
- 2.35 GB data
- HTML index structure

<table>
<thead>
<tr>
<th>Disk</th>
<th>Data Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk1</td>
<td>379,883,861</td>
</tr>
<tr>
<td>Disk2</td>
<td>473,145,614</td>
</tr>
<tr>
<td>Disk3</td>
<td>598,457,069</td>
</tr>
<tr>
<td>Disk4</td>
<td>403,880,544</td>
</tr>
<tr>
<td>Disk5</td>
<td>235,887,493</td>
</tr>
<tr>
<td>Disk6</td>
<td>258,848,527</td>
</tr>
<tr>
<td>Total</td>
<td>2,350,103,108</td>
</tr>
</tbody>
</table>
A Sample

“Hoed corn and went to ...”
Reducing Color to Grayscale

1.3 GB (23% reduction)
Reducing JPEG Quality/Color Space

629 MB (63% reduction)
Reducing Color and Quality

591 MB (66% reduction)
**DVD Media**

- **DVD-ROM**
  - High ramp-up costs to produce a run
  - Moderate market penetration compared with CD
- **Recordable DVD**
  - DVD-R, DVD-RAM, DVD-RW, DVD+RW, DVD+R
  - Incompatibility issues
  - Second generation drives are just now being released
Managing Larger Collections

Typical project: 2000 CD’s
Possibilities:
- Distribute on roughly 100 DVD’s
- Provide Web-based access
- Distribute index CD or DVD with links to online full-resolution repository

In any case, full-resolution online repository might be desirable
Storage Requirements

2000 CD’s
x 60% typical capacity
x 680 MB
= 816 Gigabytes

Can be purchased off the shelf for less than $15,000
Bandwidth Challenges

- In spite of huge investments, bandwidth is still an issue
  
  - My North Orem neighborhood, for example, is not economically served by broadband

- Great opportunity for multi-resolution browsing techniques
Transcripts

- Text has excellent properties
  - Easier to index and search
  - Lower bandwidth requirements
- Could solicit volunteers to help transcribe
- Offer incentives:
  - Training
  - Free access to the repository
  - Fame
Summary

- Quantity of historical documents is practically unlimited
- A variety of digital library technologies and strategies will be required
- Most of the technology we need is already available
- Strategies and business cases need to be worked out