Catholic, Collaborative, and Crowdfunded: A Unique Approach to Newspapers Digitization

Patricia Lawton, Catholic Research Resources Alliance
Jeff Moyer, REVEAL DIGITAL
Frederick Zarndt, Digital Divide Data

Catholic, Collaborative, and Crowdfunded: A Unique Approach to Newspapers Digitization
The Project

Patricia Lawton
Frederick Zarndt
Building the Catholic News Archive

A project of the CRRA Catholic Newspapers Program to
provide enduring access to extant Catholic newspapers in
North America
About CRRA

• Mission to provide enduring global access to Catholic research resources
• Nonprofit membership organization
• Primary projects: the Catholic Portal and the Catholic Newspapers Program
• Highly collaborative organization, with just 2 FT staff
• Open access, collaboration, and service are core values
1. Digitize 1.5 million pages from priority papers ($2.6M)
   (9 diocesan papers from major US cities, 2 national, making possible comparative analyses)

2. Aggregate content in the Veridian platform

3. Preserve content in MetaArchive, a dark archive

Reality (costs) and ideology (open access) are reconciled through a path to open access (crowdfunding model) and identifying project phases
Pilots and Phases

Samples, a slice of, the larger project, involving multiple institutions holding a mix of source material in print and microfilm.

- Civil War years
- European Immigration (1880-1919)
- Pre-1924
- Vatican II years (1958-1972)
Vatican II Timeline

### CRRA Vatican II Project Timeline

<table>
<thead>
<tr>
<th>Key Activities</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>JAN</td>
<td>FEB</td>
<td>MAR</td>
</tr>
<tr>
<td>Project start with Board approval</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complete digitization agreements</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obtain copyright permissions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify sources for materials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-call to CRRA for funding</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public call to non-CRRA libraries</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Full funding achieved</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scanning and conversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platform development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta release (funders exclusive)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary report</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project finished</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Open access begins 10/1/2017*
CRRA Technical Specifications

• Foundation is Library of Congress National Digital Newspaper Program as a foundation
• Influenced by Beyond NDNP Working Group
• Article layout influenced by National Library of Australia’s Trove newspapers and Boston University digital newspaper collection
• CRRA specifications will be made public
CRRA Workflow

1. Use rights collected by CRRA
2. Newspapers or microfilm dispatched by CRRA from source repository to scanning service bureau
3. TIFF images sent on hard drives to Digital Divide Data’s SE Asia production offices
4. Output data – METS/ALTO XML, TIFF and JPEG2000 images, PDF files – sent from SE Asia to DL Consulting (Hamilton, New Zealand)
5. DL Consulting ingests output data into Amazon AWS and EC2 based Veridian
6. After ingest, output data forwarded to CRRA for deposit in dark archive (MetaArchive)
Collaboration

Patricia Lawton
Collaboration

African proverb:
If you want to go fast, go alone.
If you want to go far, go together.

CRRA Groups

The CRRA is a distributed and collaborative organization comprised of individuals from member institutions working together in a variety of ways to support and advance the vision, mission, and activities of the CRRA.

Leadership Groups

- CRRA Board of Directors
- The Leadership Council
- The Liaisons Council
- Official Representatives Council

Committees

- Catholic Newspapers Committee
  - CRRA Digitizing Partners
  - Digitizing Agreement Group (work completed)
  - Newspapers Digitization Specifications Working Group (work completed)
  - Repository Directory Group (work completed)
- Collections Committee
  - Subject Guide Subcommittee
- Digital Access Committee
  - Metadata Subcommittee
- Membership Committee
- Scholars Advisory Committee
- Five Year Strategic Planning Committee (completed)
Where is the source material and how much will the project cost?

30 “Digitizing Partners” provide essential support in identifying where the source material lives, counting pages to determine cost estimates, and serving as liaisons to copyright holders.
Vatican II Collaborators

11 “Sourcing Partners”

11 Copyright holders

4 Vendors

- Digital Divide Data (conversion and specs)
- DL Consulting (Veridian platform)
- Master Enterprises, Inc. (image capture)
- Reveal Digital (cost recovery and project support)
Digitization agreements

1. CRRA as a primary agent in fostering digitization, dissemination, and discoverability of digitized content
2. The copyright owners
3. The institutions providing the content to be digitized (Sourcing Partners)
Crowd funding and cost recovery

Jeff Moyer
REVEAL DIGITAL
Library Crowdfunding

1. Working with Reveal Digital’s unique cost recovery = open access funding model
2. Based on the assumption that libraries will invest in open access projects
3. Reveal Digital’s initial project, *Independent Voices*, has raised over $1.2M from 78 libraries toward building an open access collection of alternative press periodicals
Step 1: Define Project Costs

Total Project Cost: $419,644

- Conversion: $216,057
- Systems: $68,875
- Outreach: $52,125
- Digital Archiving: $37,255
- Admin: $12,332
- Proj. Coord.: $33,000
Step 2: Set Funding Levels

<table>
<thead>
<tr>
<th>Institution Type</th>
<th>Target Count of Funding Libraries</th>
<th>One-time Commitment Level</th>
<th>Target Funding Raised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diocesan/Religious</td>
<td>10</td>
<td>$200</td>
<td>$2,000</td>
</tr>
<tr>
<td>Theological / Special</td>
<td>23</td>
<td>$500</td>
<td>$11,500</td>
</tr>
<tr>
<td>Bachelors</td>
<td>26</td>
<td>$3,400</td>
<td>$88,400</td>
</tr>
<tr>
<td>Masters</td>
<td>27</td>
<td>$4,250</td>
<td>$114,750</td>
</tr>
<tr>
<td>Doctorate</td>
<td>13</td>
<td>$5,100</td>
<td>$66,300</td>
</tr>
<tr>
<td>ARL</td>
<td>12</td>
<td>$8,500</td>
<td>$102,000</td>
</tr>
<tr>
<td>Large Publics</td>
<td>8</td>
<td>$4,250</td>
<td>$34,000</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>119</strong></td>
<td><strong>$4,250</strong></td>
<td><strong>$418,950</strong></td>
</tr>
</tbody>
</table>

Provides development funding for conversion, platform development plus three years of hosting, and archiving
Step 3: Library Outreach

1. Initial funding call to CRRA members
   – 20 funding libraries / $56,000 raised
2. Next funding call will target non-CRRA Catholic Institutions and other religious libraries
3. Final funding call will be made to remaining academic and large public libraries
4. Will work with consortia and other library member organizations on the funding calls
Benefits to investing institutions

1. Early access to the collection while it is being built and for a 6-month grace period
2. Free title-level MARC records for local loading and for sharing with discovery services
3. Support for mass downloading of text for patron use
4. Ensuring that Catholic newspapers become an accessible archive for all
Long term strategy for access

• Access and archiving costs included for 3 years in the initial cost model
• Explore hosting, access /delivery and archiving/preservation requirements and options
• Understand stakeholder interests
• Explore opportunities with potential partners
• Identify funding strategies and models
• Develop a sustainability plan
Patricia Lawton, Catholic Research Resources Alliance

Jeff Moyer, REVEAL DIGITAL

Frederick Zarndt, Digital Divide Data

Questions?