Using the Internet Archive and Theological Commons as an On-Ramp to Digitization and Access for Local Collections

ATLA 2017
gregory.murray@ptsem.edu
Premise
Premise

Theology libraries have important undigitized collections, unique materials
Premise

Theology libraries have important undigitized collections, unique materials

Public domain (or copyright-cleared)
Theology libraries have important undigitized collections, unique materials

Public domain (or copyright-cleared)

Would like to provide online access, openly
Premise

Theology libraries have important undigitized collections, unique materials

Public domain (or copyright-cleared)

Would like to provide online access, openly

Don’t have in-house equipment/staff for digitization
Premise

Theology libraries have important undigitized collections, unique materials

Public domain (or copyright-cleared)

Would like to provide online access, openly

Don’t have in-house equipment/staff for digitization

Don’t have in-house digital platform to provide online access
Premise

Theology libraries have important undigitized collections, unique materials

Public domain (or copyright-cleared)

Would like to provide online access, openly

Don’t have in-house equipment/staff for digitization

Don’t have in-house digital platform to provide online access

One option: Internet Archive + Theological Commons = digitization and access for local collections
Internet Archive
Internet Archive

Non-profit organization, values/mission: “Universal Access to All Knowledge”
Internet Archive

Non-profit organization, values/mission: “Universal Access to All Knowledge”

Founded 1996, best known for archiving the Web (Wayback Machine)
Internet Archive

Non-profit organization, values/mission: “Universal Access to All Knowledge”

Founded 1996, best known for archiving the Web (Wayback Machine)

In 2006, also began operating scanning centers across US and internationally
Internet Archive

Non-profit organization, values/mission: “Universal Access to All Knowledge”

Founded 1996, best known for archiving the Web (Wayback Machine)

In 2006, also began operating scanning centers across US and internationally

- Digitize books, manuscripts, photographs (not negatives), microforms (Indiana only), etc.
Internet Archive

Non-profit organization, values/mission: “Universal Access to All Knowledge”

Founded 1996, best known for archiving the Web (Wayback Machine)

In 2006, also began operating scanning centers across US and internationally

   Digitize books, manuscripts, photographs (not negatives), microforms (Indiana only), etc.

   From hundreds of research libraries/archives
Each library pays for digitization
Internet Archive

Each library pays for digitization

Digital image of each page, OCR full text for printed/typewritten materials
Internet Archive

Each library pays for digitization

Digital image of each page, OCR full text for printed/typewritten materials

Cost includes storage in perpetuity, online discovery/access at archive.org
Using the Internet Archive and Theological Commons as an On-Ramp to Digitization and Access for Local Collections

Premise

Internet Archive
Internet Archive
Costs
Internet Archive: Costs

$3 setup fee per “item”
Internet Archive: Costs

$3 setup fee per “item”
Typically “item” = bound volume
Internet Archive: Costs

$3 setup fee per “item”
   Typically “item” = bound volume
   Bound periodicals: volume vs. issue
Internet Archive: Costs

$3 setup fee per “item”
  Typically “item” = bound volume
  Bound periodicals: volume vs. issue
  Archival materials: folder vs. manuscript
Internet Archive: Costs

$3 setup fee per “item”

10¢ per page for standard size volumes (max. 11.5" x 15" and 4" thick)
Internet Archive: Costs

$3 setup fee per “item”

10¢ per page for standard size volumes (max. 11.5" x 15" and 4" thick)

25¢ per page for oversize/folio volumes or loose sheets (even if standard size)
Internet Archive: Costs

$3 setup fee per “item”

10¢ per page for standard size volumes (max. 11.5" x 15" and 4" thick)

25¢ per page for oversize/folio volumes or loose sheets (even if standard size)

$2 per foldout (any size)
Internet Archive: Costs

$3 setup fee per “item”

10¢ per page for standard size volumes (max. 11.5" x 15" and 4" thick)

25¢ per page for oversize/folio volumes or loose sheets (even if standard size)

$2 per foldout (any size)

Archival materials, loose photographs, newspapers: consult with IA; will not exceed 60¢ per image
Internet Archive: Costs

$3 setup fee per “item”

10¢ per page for standard size volumes (max. 11.5” x 15” and 4” thick)

25¢ per page for oversize/folio volumes or loose sheets (even if standard size)

$2 per foldout (any size)

Archival materials, loose photographs, newspapers: consult with IA; will not exceed 60¢ per image

Typical example: 300-page, standard size book: (300 * .10 = $30) + $3 setup = $33
Using the Internet Archive and Theological Commons as an On-Ramp to Digitization and Access for Local Collections

<table>
<thead>
<tr>
<th>Premise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet Archive</td>
</tr>
<tr>
<td>Costs</td>
</tr>
</tbody>
</table>
Internet Archive
Metadata
Internet Archive: Metadata

Excel file (saved as Comma Separated Values), one row per item
Internet Archive: Metadata

Excel file (saved as Comma Separated Values), one row per item

Minimum: title, language
Internet Archive: Metadata

Excel file (saved as Comma Separated Values), one row per item

Minimum: title, language

Various predefined fields (IA has a template), including “description”
Internet Archive: Metadata

Excel file (saved as Comma Separated Values), one row per item

Minimum: title, language

Various predefined fields (IA has a template), including “description”

If you provide OCLC number (or ISBN), IA can utilize WorldCat record
Internet Archive: Metadata

Excel file (saved as Comma Separated Values), one row per item

- Minimum: title, language
- Various predefined fields (IA has a template), including “description”
- If you provide OCLC number (or ISBN), IA can utilize WorldCat record

Alternative: your MARC records — if 500+ items
Internet Archive: Metadata

Excel file (saved as Comma Separated Values), one row per item

Minimum: title, language

Various predefined fields (IA has a template), including “description”

If you provide OCLC number (or ISBN), IA can utilize WorldCat record

Alternative: your MARC records — if 500+ items

Ask IA to set up Z39.50 connection to your catalog
Internet Archive: Metadata

Excel file (saved as Comma Separated Values), one row per item

Minimum: title, language

Various predefined fields (IA has a template), including “description”

If you provide OCLC number (or ISBN), IA can utilize WorldCat record

Alternative: your MARC records — if 500+ items

Ask IA to set up Z39.50 connection to your catalog

You provide bib record IDs; IA harvests MARC records, derives metadata
Using the Internet Archive and Theological Commons as an On-Ramp to Digitization and Access for Local Collections

<table>
<thead>
<tr>
<th>Premise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet Archive</td>
</tr>
<tr>
<td>Costs</td>
</tr>
<tr>
<td>Metadata</td>
</tr>
</tbody>
</table>
Theological Commons
Theological Commons

PTSem partnered with IA in early 2008 for large-scale book digitization
Theological Commons

PTSem partnered with IA in early 2008 for large-scale book digitization

Established IA scanning center in PTSem Library (equipped, staffed, operated by IA)
Theological Commons

PTSem partnered with IA in early 2008 for large-scale book digitization

Established IA scanning center in PTSem Library (equipped, staffed, operated by IA)

Partnership goals:

  Provide free, worldwide access to PTSem public domain books
Theological Commons

PTSem partnered with IA in early 2008 for large-scale book digitization

Established IA scanning center in PTSem Library (equipped, staffed, operated by IA)

Partnership goals:

  Provide free, worldwide access to PTSem public domain books

  Contribute to representation of theology and related fields within a mass-scale, global, online digital library
Theological Commons

By 2011, PTSem + IA digitized 23,000 books, available online at archive.org
Theological Commons

By 2011, PTSem + IA digitized 23,000 books, available online at archive.org

Fulfilling original goals, but ...
Theological Commons

By 2011, PTSem + IA digitized 23,000 books, available online at archive.org

Fulfilling original goals, but ...

In context of several million other texts (not to mention audio, video, images)
Theological Commons

By 2011, PTSem + IA digitized 23,000 books, available online at archive.org

Fulfilling original goals, but ...

In context of several million other texts (not to mention audio, video, images)

On every subject
Theological Commons

By 2011, PTSem + IA digitized 23,000 books, available online at archive.org

Fulfilling original goals, but ...

In context of several million other texts (not to mention audio, video, images)

On every subject

Both academic and popular
Theological Commons

By 2011, PTSem + IA digitized 23,000 books, available online at archive.org

Fulfilling original goals, but ...

In context of several million other texts (not to mention audio, video, images)

On every subject

Both academic and popular

New goal branched out:
Theological Commons

By 2011, PTSem + IA digitized 23,000 books, available online at archive.org

Fulfilling original goals, but ...

In context of several million other texts (not to mention audio, video, images)

On every subject

Both academic and popular

New goal branched out:

In addition to digitization/access as an end in itself
Theological Commons

By 2011, PTSem + IA digitized 23,000 books, available online at archive.org

Fulfilling original goals, but ...

In context of several million other texts (not to mention audio, video, images)

On every subject

Both academic and popular

New goal branched out:

In addition to digitization/access as an end in itself
provide alternative, focused discovery for content in theology/related fields
Theological Commons

By 2011, PTSem + IA digitized 23,000 books, available online at archive.org

Fulfilling original goals, but ...

In context of several million other texts (not to mention audio, video, images)

On every subject

Both academic and popular

New goal branched out:

In addition to digitization/access as an end in itself
provide alternative, focused discovery for content in theology/related fields
to serve a specialized audience
Theological Commons

1. Select content from IA based on subject
Theological Commons

1. Select content from IA based on subject
2. Build a web interface for search/navigation of that subset
Theological Commons

1. Select content from IA based on subject
2. Build a web interface for search/navigation of that subset

Publicly released in March 2012 as “Theological Commons” (commons.ptsem.edu)
Theological Commons

1. Select content from IA based on subject
2. Build a web interface for search/navigation of that subset

Publicly released in March 2012 as “Theological Commons” (commons.ptsem.edu)

Spearheaded by:
Theological Commons

1. Select content from IA based on subject
2. Build a web interface for search/navigation of that subset

Publicly released in March 2012 as “Theological Commons” (commons.ptsem.edu)

Spearheaded by:
   Cliff Anderson
Theological Commons

1. Select content from IA based on subject
2. Build a web interface for search/navigation of that subset

Publicly released in March 2012 as “Theological Commons” (commons.ptsem.edu)

Spearheaded by:
   Cliff Anderson
   Steve Crocco
Theological Commons

1. Select content from IA based on subject
2. Build a web interface for search/navigation of that subset

Publicly released in March 2012 as “Theological Commons” (commons.ptsem.edu)

Spearheaded by:
   Cliff Anderson
   Steve Crocco
   with support of Iain Torrance
Using the Internet Archive and Theological Commons as an On-Ramp to Digitization and Access for Local Collections

<table>
<thead>
<tr>
<th>Premise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet Archive</td>
</tr>
<tr>
<td>Costs</td>
</tr>
<tr>
<td>Metadata</td>
</tr>
<tr>
<td>Theological Commons</td>
</tr>
</tbody>
</table>
Theological Commons
Content
Theological Commons: Content

Selection:
PTSem content, downloaded/imported from IA
Theological Commons: Content

Selection:
  PTSem content, downloaded/imported from IA
  Imported content from other libraries, based on LCSH (over 150 libraries)
Theological Commons: Content

Selection:
PTSem content, downloaded/imported from IA
Imported content from other libraries, based on LCSH (over 150 libraries)

Attribution:
Browse by Contributor
Theological Commons: Content

Selection:
PTSem content, downloaded/imported from IA
Imported content from other libraries, based on LCSH (over 150 libraries)

Attribution:
Browse by Contributor
Item-level page
Theological Commons: Content

Selection:
- PTSem content, downloaded/imported from IA
- Imported content from other libraries, based on LCSH (over 150 libraries)

Attribution:
- Browse by Contributor
- Item-level page
- FAQ
Theological Commons: Content

Selection:
- PTSem content, downloaded/imported from IA
- Imported content from other libraries, based on LCSH (over 150 libraries)

Attribution:
- Browse by Contributor
- Item-level page
- FAQ

Formats:
- Books, periodicals, theses, mss, photos, postcards, maps (digitized via IA)
Theological Commons: Content

Selection:
- PTSem content, downloaded/imported from IA
- Imported content from other libraries, based on LCSH (over 150 libraries)

Attribution:
- Browse by Contributor
- Item-level page
- FAQ

Formats:
- Books, periodicals, theses, mss, photos, postcards, maps (digitized via IA)
- Audio/video — not digitized by IA, currently only PTSem
Using the Internet Archive and Theological Commons as an On-Ramp to Digitization and Access for Local Collections

<table>
<thead>
<tr>
<th>Premise</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet Archive</td>
<td></td>
</tr>
<tr>
<td>Costs</td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td></td>
</tr>
<tr>
<td>Theological Commons</td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td></td>
</tr>
</tbody>
</table>

ATLA 2017

gregory.murray@ptsem.edu
Theological Commons
Features
Theological Commons: Features

Mobile-friendly (Responsive Web Design, not a separate mobile version)
Theological Commons: Features

Mobile-friendly (Responsive Web Design, not a separate mobile version)

Full-text + metadata searching for texts (books, periodicals, theses, typewritten manuscripts, audio/video transcripts)
Theological Commons: Features

Mobile-friendly (Responsive Web Design, not a separate mobile version)

Full-text + metadata searching for texts (books, periodicals, theses, typewritten manuscripts, audio/video transcripts)

Metadata only searching for media (photographs, postcards, audio/video without transcripts)
Theological Commons: Features

Mobile-friendly (Responsive Web Design, not a separate mobile version)

Full-text + metadata searching for texts (books, periodicals, theses, typewritten manuscripts, audio/video transcripts)

Metadata only searching for media (photographs, postcards, audio/video without transcripts)

Facets for filtering search results
Theological Commons: Features

Mobile-friendly (Responsive Web Design, not a separate mobile version)

Full-text + metadata searching for texts (books, periodicals, theses, typewritten manuscripts, audio/video transcripts)

Metadata only searching for media (photographs, postcards, audio/video without transcripts)

Facets for filtering search results
  Allow selection of multiple values simultaneously
Theological Commons: Features

Mobile-friendly (Responsive Web Design, not a separate mobile version)

Full-text + metadata searching for texts (books, periodicals, theses, typewritten manuscripts, audio/video transcripts)

Metadata only searching for media (photographs, postcards, audio/video without transcripts)

Facets for filtering search results
   Allow selection of multiple values simultaneously
   Also provide browsing
Theological Commons: Features

Mobile-friendly (Responsive Web Design, not a separate mobile version)

Full-text + metadata searching for texts (books, periodicals, theses, typewritten manuscripts, audio/video transcripts)

Metadata only searching for media (photographs, postcards, audio/video without transcripts)

Facets for filtering search results
  - Allow selection of multiple values simultaneously
  - Also provide browsing

Read online at archive.org
Using the Internet Archive and Theological Commons as an On-Ramp to Digitization and Access for Local Collections

<table>
<thead>
<tr>
<th>Premise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet Archive</td>
</tr>
<tr>
<td>Costs</td>
</tr>
<tr>
<td>Metadata</td>
</tr>
<tr>
<td>Theological Commons</td>
</tr>
<tr>
<td>Content</td>
</tr>
<tr>
<td>Features</td>
</tr>
</tbody>
</table>

gregory.murray@ptsem.edu
Theological Commons
Collaborations
Theological Commons: Collaborations

Established platform, workflow for importing content from other libraries
Theological Commons: Collaborations

Established platform, workflow for importing content from other libraries

Set up Featured Collections
Theological Commons: Collaborations

Established platform, workflow for importing content from other libraries

Set up Featured Collections

Payne Theological Seminary and A.M.E. Church Archive
Theological Commons: Collaborations

Established platform, workflow for importing content from other libraries

Set up Featured Collections

Payne Theological Seminary and A.M.E. Church Archive
Small project: digitized 86 items, incorporated 56 items from other libraries
Theological Commons: Collaborations

Established platform, workflow for importing content from other libraries

Set up Featured Collections

Payne Theological Seminary and A.M.E. Church Archive
Small project: digitized 86 items, incorporated 56 items from other libraries

Missionary Research Library Pamphlets (Burke/Union/Columbia)
Theological Commons: Collaborations

Established platform, workflow for importing content from other libraries

Set up Featured Collections

Payne Theological Seminary and A.M.E. Church Archive
   Small project: digitized 86 items, incorporated 56 items digitized elsewhere

Missionary Research Library Pamphlets (Burke/Union/Columbia)
   Ongoing
Using the Internet Archive and Theological Commons as an On-Ramp to Digitization and Access for Local Collections

<table>
<thead>
<tr>
<th>Premise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet Archive</td>
</tr>
<tr>
<td>Costs</td>
</tr>
<tr>
<td>Metadata</td>
</tr>
<tr>
<td>Theological Commons</td>
</tr>
<tr>
<td>Content</td>
</tr>
<tr>
<td>Features</td>
</tr>
<tr>
<td>Collaborations</td>
</tr>
</tbody>
</table>
Premise

Theology libraries have important undigitized collections, unique materials

Public domain (or copyright-cleared)

Would like to provide online access, openly

Don’t have in-house equipment/staff for digitization

Don’t have in-house digital platform to provide online access

One option: Internet Archive + Theological Commons = digitization and access for local collections
IA + TC + Your Library
IA + TC + Your Library

Digitization workflow exists (IA), platform exists (TC)
IA + TC + Your Library

Digitization workflow exists (IA), platform exists (TC)
Platform includes Featured Collections
IA + TC + Your Library

Digitization workflow exists (IA), platform exists (TC)
Platform includes Featured Collections
Established model for other libraries to have their own featured collections
IA + TC + Your Library

Digitization workflow exists (IA), platform exists (TC)
Platform includes Featured Collections
Established model for other libraries to have their own featured collections

Opportunity for any library to bring its materials online via same model
IA + TC + Your Library

Digitization workflow exists (IA), platform exists (TC)
Platform includes Featured Collections
Established model for other libraries to have their own featured collections

Opportunity for any library to bring its materials online via same model

Digitize via IA
IA + TC + Your Library

Digitization workflow exists (IA), platform exists (TC)
   Platform includes Featured Collections
   Established model for other libraries to have their own featured collections

Opportunity for any library to bring its materials online via same model

Digitize via IA
   Additional metadata we request:
      Identify any multi-volume sets (powers “View All Volumes” links)
Digitization workflow exists (IA), platform exists (TC)

Platform includes Featured Collections
Established model for other libraries to have their own featured collections

Opportunity for any library to bring its materials online via same model

Digitize via IA

Additional metadata we request:
Identify any multi-volume sets (powers “View All Volumes” links)
Subject term from our list, or LC call number (powers Subject facet)
IA + TC + Your Library

Digitization workflow exists (IA), platform exists (TC)
  Platform includes Featured Collections
  Established model for other libraries to have their own featured collections

Opportunity for any library to bring its materials online via same model

Digitize via IA
  Additional metadata we request:
    Identify any multi-volume sets (powers “View All Volumes” links)
    Subject term from our list, or LC call number (powers Subject facet)

PTSem Library sets up featured collection, imports your content (at no cost)
IA + TC + Your Library

Digitization workflow exists (IA), platform exists (TC)
   Platform includes Featured Collections
   Established model for other libraries to have their own featured collections

Opportunity for any library to bring its materials online via same model

Digitize via IA
   Additional metadata we request:
      Identify any multi-volume sets (powers “View All Volumes” links)
      Subject term from our list, or LC call number (powers Subject facet)

PTSem Library sets up featured collection, imports your content (at no cost)
   No technology staff needed
IA + TC + Your Library

No added cost to contribute content to TC
IA + TC + Your Library

No added cost to contribute content to TC

Your materials enhance the content of TC
IA + TC + Your Library

No added cost to contribute content to TC

Your materials enhance the content of TC

A service to theological community
IA + TC + Your Library

No added cost to contribute content to TC
Your materials enhance the content of TC
A service to theological community

Necessity of grant funding for digitization
IA + TC + Your Library

No added cost to contribute content to TC

Your materials enhance the content of TC

A service to theological community

Necessity of grant funding for digitization

Key to focus on unique materials
IA + TC + Your Library

No added cost to contribute content to TC

Your materials enhance the content of TC

A service to theological community

Necessity of grant funding for digitization

Key to focus on unique materials

IA is one of eight finalists for MacArthur Foundation’s “100 & Change” grant
IA + TC + Your Library

No added cost to contribute content to TC
Your materials enhance the content of TC
A service to theological community

Necessity of grant funding for digitization
Key to focus on unique materials

IA is one of eight finalists for MacArthur Foundation’s “100 & Change” grant
If awarded, libraries would apply to IA directly for digitization funding
IA + TC + Your Library

This model provides two complementary means of discovery/access:
IA + TC + Your Library

This model provides two complementary means of discovery/access:

IA: wide reach, in context mass digitization online library, millions of users
IA + TC + Your Library

This model provides two complementary means of discovery/access:

IA: wide reach, in context mass digitization online library, millions of users

IA will feed your content into DPLA if you authorize that
IA + TC + Your Library

This model provides two complementary means of discovery/access:

IA: wide reach, in context mass digitization online library, millions of users

IA will feed your content into DPLA if you authorize that

TC: focused, specialized for targeted audience
IA + TC + Your Library

This model provides two complementary means of discovery/access:

IA: wide reach, in context mass digitization online library, millions of users

IA will feed your content into DPLA if you authorize that

TC: focused, specialized for targeted audience

Both serve theological community
IA + TC + Your Library

This model provides two complementary means of discovery/access:

IA: wide reach, in context mass digitization online library, millions of users
IA will feed your content into DPLA if you authorize that

TC: focused, specialized for targeted audience

Both serve theological community
Both support core mission of our libraries
Using the Internet Archive and Theological Commons as an On-Ramp to Digitization and Access for Local Collections

ATLA 2017

gregory.murray@ptsem.edu