

Designing & Maintaining an EIS Database:

Lessons Learned in Developing Library-based Digital Repositories

Robert Cartolano

Associate VP, Digital Program & Technology Services

June 16, 2016

Topics

- Overview of Libraries
- Six Lessons Learned
 - Find Partners and Collaborate
 - Listen to Your Users
 - Preservation and Access
 - Decouple Front-end from Back-end
 - Metadata, metadata, metadata
 - Innovate at the Front-end
- Q & A



Overview of Libraries

- 21 Libraries
- 12 million volumes
- 160,000 journals and serials
- 450 professional and support staff
- 4,000+ study seats
- 1,100 workstations
- Significant technology infrastructure
- Thousands of external vendors
- Over \$10M annual spend on E-resources

<http://library.columbia.edu/about.html>



New Type of Research Library

“While the core features traditionally associated with research libraries continue to play a central role, our services embrace a broader range, extending the conception of the research library.”

- James Neal, Former VP for Information Services and University Librarian



1. Find Partners, Collaborate

- Internal and external partners
- Partners share development efforts
- Leverage open source efforts
- Communities support each other
- Communities can help obtain external funding for targeted projects



Academic Preservation Trust (APTrust)

- University of Virginia, 17 partner institutions
- APTrust is working toward TRAC certification
- Repository to be built from the ground up taking TRAC certification into account.
- Certification Working Group established and will be advising and consulting with the APTrust staff and partners on TRAC objectives.
- Initial development work is proceeding at the level of Digital Object Management and Infrastructure.



<http://aptrust.org/>



External Partners

- **Blacklight**

- discovery software
- <http://projectblacklight.org/>



- **Fedora**

- digital preservation repository software
- <http://www.fedora-commons.org/>



- **Hydra**

- digital content management
- <http://projecthydra.org/>



- **DuraSpace**

- 138 member institutions
- <http://duraspace.org/>

A screenshot of a website page. At the top, there is a dark blue navigation bar with white text links: "Project", "Get Started", "Software", "Community", "Membership", and "Resources". Below the navigation bar is a testimonial section. On the left is a portrait of Robert Cartolano, a man in a suit and tie. To the right of the portrait is a quote: "Fedora is the foundation of Columbia's digital preservation storage system. We chose Fedora in order to preserve and make accessible our digital collections because we can leverage the flexibility of the system for a wide variety of uses, and because we can work...". Below the quote is a link that says "READ MORE".

Project Get Started Software Community Membership Resources

Robert Cartolano
Associate Vice President, Digital Programs and
Columbia University

"Fedora is the foundation of Columbia's digital preservation storage system. We chose Fedora in order to preserve and make accessible our digital collections because we can leverage the flexibility of the system for a wide variety of uses, and because we can work..."

[READ MORE](#)

2. Listen to Your Users

- Gather data - Assessments, surveys, focus groups, needs analysis
- Real-world user testing - capture activity, video session
- Engage users in any way you can



3. Preservation and Access

Preservation - prolong the **existence** and **accessibility** of the collections for current and future students and scholars

<http://library.columbia.edu/services/preservation.html>

Backup ≠ Preservation



Digital Library Collections

We must preserve and provide access to ...

- Local Digitization Projects
- Preservation-Related Digitization
- Institutional Repository / Data Sets
- Born Digital Archival Content
- Archived Web Sites
- Super Dark Archives – *highly secure*





Census Data

<http://www.census.gov/data.html>



ICE Pod - Antarctic Research

<http://www.ideo.columbia.edu/res/pi/icepod/>

Ancient Ink



Manuscript fragment from the Rare Book and Manuscript Library, Columbia University, Col. Inv. 581, in Coptic and Greek on papyrus, 12 x 6 cm, Estimate#2009-00102

<http://cise.columbia.edu/AncientInk>



SEDAC

<http://sedac.ciesin.org/>



16th Century Dante Books

<http://digitaldante.columbia.edu/>



<http://www.iedadata.org/>

duPont-Columbia Awards Videos

<http://www.journalism.columbia.edu/dupont>



COLUMBIA UNIVERSITY LIBRARIES / INFORMATION SERVICES

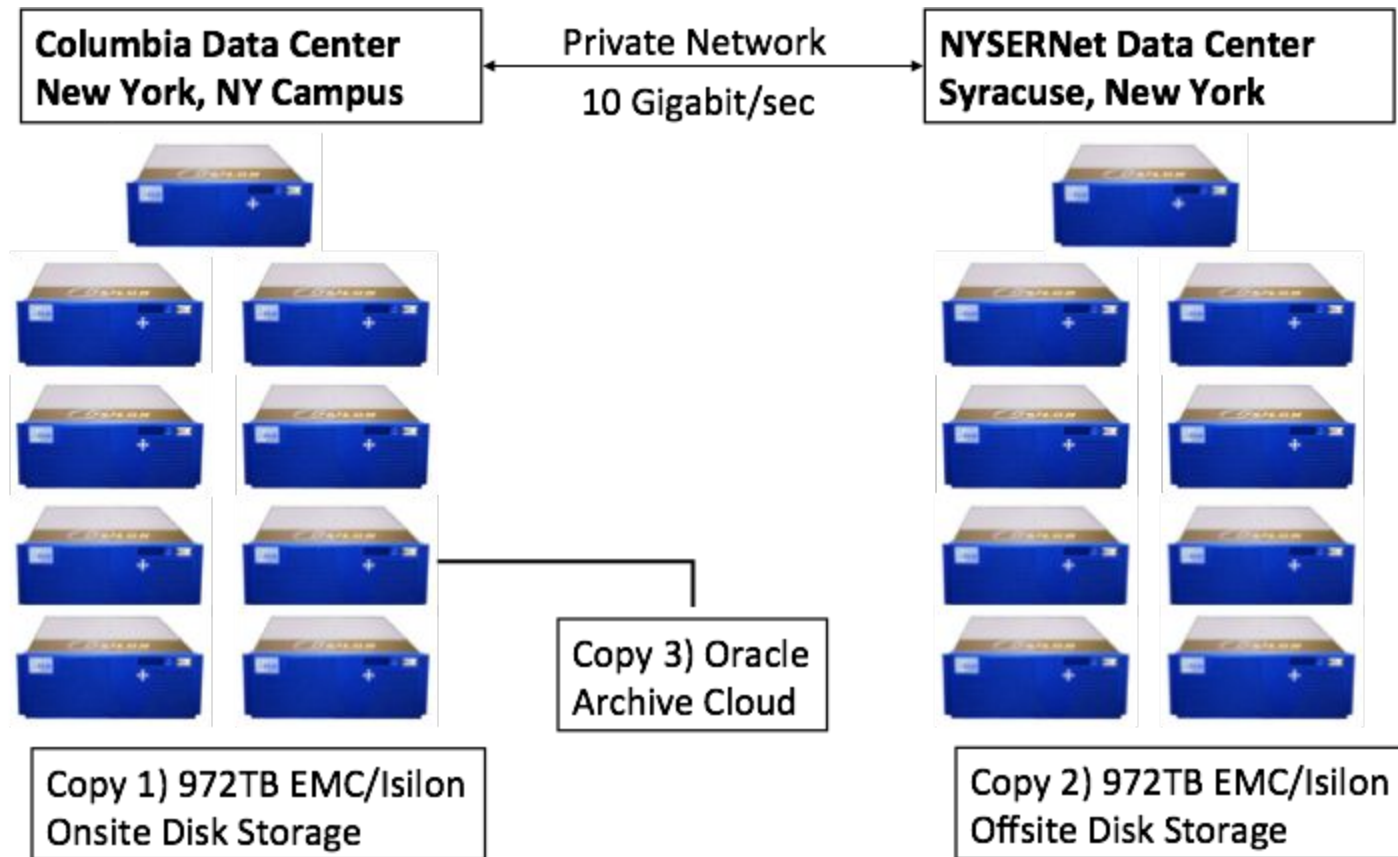
4. Decouple Front-end & Back-End

- Separation provides flexibility
 - Risk-averse storage approach
 - Different storage for preservation, access
 - Accommodate multiple front-end systems
 - Leverage national/cloud archive solutions
- Swap out back-end storage
 - 2009 - 70 Terabytes, 4 copies
 - 2013 - 300 Terabytes, 3 copies
 - 2016 - 800 Terabytes, 3 copies



2016 Library Digital Preservation Storage

- Two copies on disk, third copy archived in cloud
- 800 Terabytes effective storage (~2PB total)



Fedora Software Platform

<http://fedorarepository.org/>

Flexible **E**xtensible **D**igital **O**bject **R**epository **A**rchitecture

- Open source digital repository software
- Internationally developed, supported, adopted
- Platform for Columbia's digital preservation system

Fedora Overview: <http://dx.doi.org/10.7916/D8FF3QDR>

Columbia Profile: <http://www.fedora-commons.org/profile/101>



Key Features of Fedora 4

<http://fedorarepository.org/>

- Authentication/Authorization
- Backup
- Clustering
- Content modeling
- Fixity services
- Linked data (native RDF support)
- Search (admin and external)
- SPARQL endpoints (semantic protocol, RDF query language)
- Advanced storage capabilities
 - Policy-driven, self-healing, very large files
 - External data source “projection” (aka “virtual ingest”)
- Transactions
- Versioning



5. Metadata, Metadata, Metadata

- Citable metadata in standard formats with persistent identifiers fosters long-term access and data management
- Leverage human and computer-generated descriptive metadata for search and discovery
- Assign rights, administrative, provenance, technical, structural and preservation metadata for long-term curation



Leverage Standards

- Portland Content Data Model (PCDM)
- RDF/Linked Data for Libraries
- Text mark-up standards (XML, TEI)
- Preservation Standards (ISO 14721 - OASIS; ISO 16363 - Trustworthy Repositories)
- New and emerging web, content and technology standards



6. Innovate at the Front-End

<http://projectblacklight.org/>



Blacklight Achieves...

- Search capabilities that meet user expectations
- Speed
- Flexibility to customize
- Unified Discovery
- Vibrant community - continuous, iterative development



Blacklight Does Much More...

- Relevance ranking ✓
- Stemming ✓
- Field weighting ✓
- Spell-checking ✓
- Refining original queries ✓
- Support for popular query operators ✓
- Boolean bag o' goods ✓
- Flexible default query processing
- In-line query limiters
- Duplicate detection
- Sort Flexibility ✓
- Character Sets (Unicode UTF-8) ✓
- Faceting ✓
- Advanced search ✓
- Easily customized search results
- Best bets ✓
- Search logging and reporting ✓
- Well-rounded administrative interface



Blacklight Community

- 469 Members (328 in April 2014)
 - blacklight-development@googlegroups.com
- Over 100 Organizations (estimated)
- Part of the Hydra Project



200 HydraNauts from 60 institutions at Hydra Connect 2015
<https://wiki.duraspace.org/display/hydra/Hydra+Connect+2015>

Unified Discovery - single search box

COLUMBIA UNIVERSITY LIBRARIES

HOME ▾ LOG OUT ▾ FIND ▾ RESEARCH SUPPORT ▾ SERVICES ▾ COMPUTERS & TECHNOLOGY ▾

SEARCH ▾ (Catalog, Articles, Academic Commons, Website Search)

Go Q

Search for books, articles, journals, videos, music, and more. Use [CLIO](#) for advanced search features.



CLIO

<http://clio.columbia.edu/>

CLIO

Quicksearch

- [Catalog](#)
- [Articles](#)
- [E-Journal Titles](#)
- [Databases](#)
- [Academic Commons](#)
- [Libraries Website](#)
- [Archives](#)
- [Dissertations](#)
- [E-Books](#)
- [New Arrivals](#)
- [Newspapers](#)

Catalog View and filter all 4,691 results

Articles View and filter all 2,150,413 results

Academic Commons View and filter all 34 results

Libraries Website View all 44 results

Delta : vivre et travailler dans le delta intérieur du fleuve Niger au Mali

Author Coulon, Gilles
Published Bamako : Donniya ; Paris : IRD, c2000.
Location ✓ Lehman DT551.24 .C68 2000g
Format Book

Delta [videorecording]

Published [United States] : Facets Multimedia, [2012]
Location ✓ Butler Media (Circulating), 208 Butler DVD22564
Format Video

Delta

Published Thoubal, Manipur : Seawaves, Manipur, 2008.
Location ✗ Offsite PL4001.M317 D45 2008g
Format Book

DELTA [electronic resource]

Published [S.l.] : Tu Delft
Online ✓ Open Access Journals: 1994 to present
Format Journal/Periodical, Online

Delta : laviuri

Author Avachian, Hrandt
Published București : "Litera", 1976.
Location ✓ Avery Fine Arts (Non-Circulating) ND933 Av11 Av11 Q
Format Book

Delta : Nederlands verleden in vogelvlucht

Published Leiden : M. Nijhoff, 1992-
Location ✓ Butler Stacks DH107 .D46 1992g

delta

Citation A Dictionary of Finance and Banking, 1/1/2014
Format Reference: [Available Online](#)

Delta

Author Lazar, Alexandra
Citation Third Text, 7/1/2010, ISSN: 0952-8822, Volume 24, Issue 4 p. 498
Format Journal Article: [Full Text Available](#)

Paleocontinental Setting for the Catskill Delta

Author Kent, Dennis V.
Date 1985
Subject Geology, Geomorphology
Downloads Kent1985b.pdf

Salinity Crisis in the Borderlands: The Mexicali Campesino's Fight for Colorado River Water in the Delta, 1961-1973

Author Gottlieb, Jillian Gordon
Date 2012
Subject Environmental justice, Latin American studies
Downloads gottlieb_thesis.pdf

UTS: Sam and Dorothy Franklin Papers, 1910-1994

... Bulk dates: 1924 – 1989 Abstract: Sam Horace Franklin, Jr., missionary to Japan with wife Dorothy, director and pastor of Delta and Providence

CLIO Technologies

<https://blogs.cul.columbia.edu/cliol/>

- Blacklight/SOLR front-end (Library infra)
- Ex Libris Voyager back-end (CUIT infra)
- EZProxy proxy service for remote, authenticated access to e-resources (Library infra)
- Summon Article Search - 1.4 billion articles
- Link Resolver - 2M+ permanent URLs
- CY2015: 1.9M sessions, 9.9M page views
- Includes:
 - Integration w/Law Library collection (400,000+ items)
 - HathiTrust ebook titles



Academic Commons

<http://academiccommons.columbia.edu/>

 COLUMBIA UNIVERSITY
ACADEMIC COMMONS



Search



Academic Commons is Columbia University's digital repository where faculty, students, and staff of Columbia and its affiliate institutions can deposit the results of their scholarly work and research. Content in *Academic Commons* is freely available to the public.



Login

BROWSE CONTENT

DEPARTMENTS
SUBJECTS

HELPFUL LINKS

DEPOSIT YOUR RESEARCH
ABOUT ACADEMIC COMMONS
FAQ/ASK A QUESTION

REPOSITORY STATISTICS

14020 items in Academic Commons.
Objects added in the last year: 2546
Objects added in the last 30 days: 210
Visitors in October: 32156

New in Academic Commons:

Herve Lacombe. *The Keys to French Opera in the Nineteenth Century*. Translated by Edward Schneider. Berkeley: University of California Press, 2001. xv, 415 pp.

[Speare, Mary Jean](#)

Review of David Charlton. *French Opera 1730-1830: Meaning and Media*. Aldershot, Hampshire (U.K.), and Brookfield, Vt.: Ashgate, 2000. xii, 374 pp.

[Rushton, Julian](#)

Review of David Ferris. Schumann's "Eichendorff" Liederkreis" and the Genre of the Romantic Cycle. Oxford and New York: Oxford University Press, 2000. vii, 270 pp.

Review of Marian Smith. *Ballet and Opera in the Age of "Giselle"*. Princeton: Princeton University Press, 2001. xx, 306 pp.

Review of John S. Powell. *Music and Theatre in France, 1600-1680*. Oxford: Oxford University Press, 2000. xvii, 582 pp.

Academic Commons

<http://academiccommons.columbia.edu/>

- Columbia's repository for faculty, students
- 19,000 items
- Dissertations, journal articles, more
- Blacklight/SOLR front-end
- Fedora repository back-end



Digital Library Collections

<http://dlc.library.columbia.edu/>



DIGITAL LIBRARY COLLECTIONS

All Fields ▾

Start a new search...

Search 🔍

WELCOME

The Digital Library Collections (DLC) website is a gateway to digital reproductions and descriptions of photographs, posters, drawings, objects, ephemera, and manuscripts as well as other archival material from Columbia's rare and special collections. As of January 2016 the DLC repository includes ca. 190,000 digital objects. [More »](#)

COLLECTION SPOTLIGHT

Chinese Paper Gods

An online visual catalog of over 200 woodcuts used in folk religious practices in Beijing and



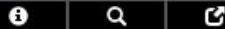
SELECTED COLLECTIONS

NAMES

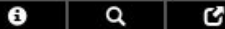
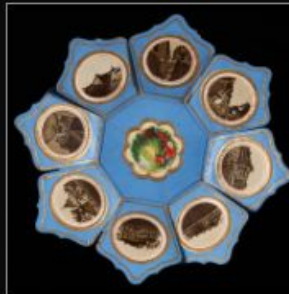
FORMATS

LIBRARY LOCATIONS

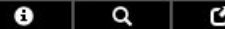
ALL COLLECTIONS



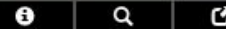
Avery's Architectural Ephemera Collections



Avery's Architectural Novelties



Biggart Collection of Architectural Vignettes on Commercial Stationery



Children's Drawings of the Spanish Civil War

Durst Old York Library

<http://library.columbia.edu/locations/avery/seymour-b-durst-old-york-library.html>

COLUMBIA UNIVERSITY LIBRARIES / INFORMATION SERVICES Help ▾

SEYMOUR B. DURST OLD YORK LIBRARY

AVERY ARCHITECTURAL & FINE ARTS LIBRARY

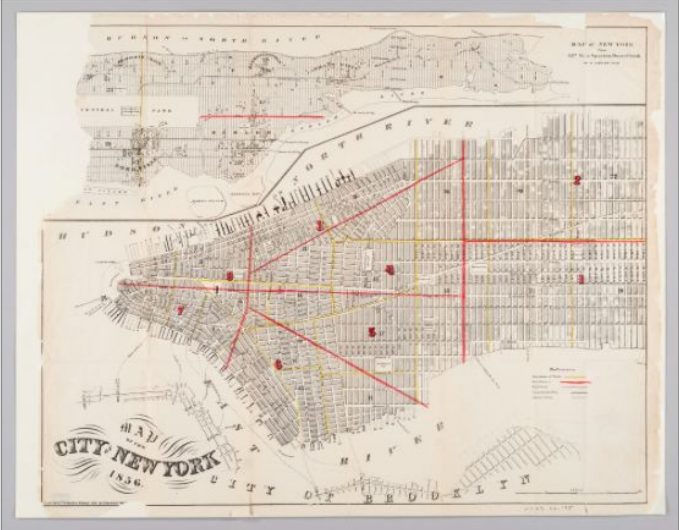
Formats ▾ Search Postcards, Maps, Photographs, Books, Etc...

LEARN MORE ▾

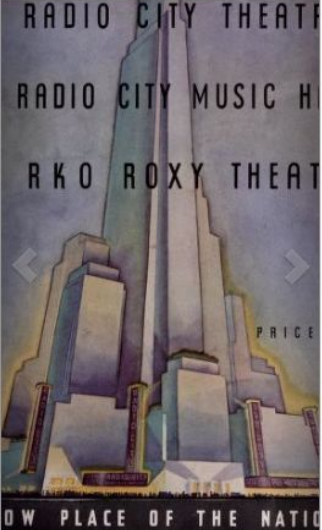
ABOUT THE OLD YORK COLLECTION

The Seymour B. Durst Old York Library collection at the Avery Architectural & Fine Arts Library consists of more than 40,000 objects including historic photographs, maps, pamphlets, postcards, books, and New York City memorabilia from the 18th century to the 1980s. [Learn More >](#)

FIND BY LOCATION



FIND IMAGES



FIND BY

FORMAT

> Postcards	14,568
> Books	8,772
> Other	597
> Periodicals	549
> Maps	449
> Ephemera	156
> Objects	15
> Manuscripts	13
> Music	6

NEIGHBORHOOD ▾

BOROUGH ▾

CITY ▾



Web Archiving

600+ sites, 50M indexed pages, 9TB, Internet Archive

<http://hrwa.cul.columbia.edu/>

The screenshot shows the homepage of the Human Rights Web Archive. At the top, there is a navigation bar with the Columbia University Libraries logo, the acronym 'HRWA', a search bar with the placeholder text 'Start a new search...', and a 'Search' button. Below the navigation bar is a large banner with the text 'HUMAN RIGHTS WEB ARCHIVE' and 'CENTER FOR HUMAN RIGHTS DOCUMENTATION & RESEARCH AT COLUMBIA UNIVERSITY'. A 'BETA' label is visible in the top right corner of the banner. Below the banner is a secondary navigation bar with links for 'LEARN MORE', 'NOMINATE A SITE', and 'CONTACT', along with a 'SEARCH HISTORY' link. The main content area features a 'FEATURED' tab and a list of featured subjects. The first featured subject is 'Transitional justice', with a selection of sites related to it. Three site cards are visible, each with a thumbnail image, a title, and a 'Site Details' button. The titles of the visible cards are 'National plan of action (NPoA) for the I...', 'Institute for Justice and Reconciliation', and 'Coalition for Women's Human Rights in...'. The left sidebar contains a description of the archive and information about its funding.

COLUMBIA UNIVERSITY LIBRARIES - HRWA Start a new search... Sites Search

HUMAN RIGHTS WEB ARCHIVE

CENTER FOR HUMAN RIGHTS DOCUMENTATION & RESEARCH AT COLUMBIA UNIVERSITY

BETA

LEARN MORE - NOMINATE A SITE - CONTACT SEARCH HISTORY

The Human Rights Web Archive @ Columbia University is a searchable collection of archived copies of human rights websites created by non-governmental organizations, national human rights institutions, tribunals and individuals. Collecting began in 2008 and has been ongoing for active websites. New websites are added to the collection regularly.

The HRWA is an initiative of the [Center for Human Rights Documentation & Research](#) and is a key focus of the Columbia University Libraries' [Web Resources Collection Program](#). The HRWA was made possible by generous support from the Andrew W. Mellon Foundation. [Learn More >](#)

FEATURED WEBSITES URLS SUBJECTS PLACES LANGUAGES

A selection of sites relating to the subject **Transitional justice**. Featured Subjects

National plan of action (NPoA) for the I... Site Details

Institute for Justice and Reconciliation Site Details

Coalition for Women's Human Rights in... Site Details

Archives Dealing with the Past Swisspeace

MLA Core Beta

Beta Site at: <https://commons.mla.org/core/>

Press Release: <https://commons.mla.org/core/press-release/>



[GROUPS](#) [MEMBERS](#) [SITES](#) [ACTIVITY](#) [CORE](#) [PUBLICATIONS](#) [HELP](#)

Browse by Subject

- [Academe](#) | 9
- [Aesthetics](#) | 4
- [African American culture](#) | 2
- [African American literature](#) | 3

[more>>](#)

Browse by Item Type

- [Article](#) | 55
- [Bibliography](#) | 2
- [Blog Post](#) | 1
- [Book](#) | 3

[more>>](#)

Browse by Date

CORE Deposits Directory

All Deposits 115

Order By:

[Local Rooms with a Cosmopolitan View? Novels in/on the Limits of European Convergence](#)

Author(s): [Cesar Dominguez \(see profile\)](#)

Date: 2015

Subject(s): [Comparative literature](#), [European literature](#)

Permanent URL: <http://dx.doi.org/10.17613/M62C76>

[Post-Traumatic Responses in the War Narratives of Hanan al-](#)

Author(s): [Ghada Mohammad](#), [Majda R. Atieh \(see profile\)](#)

Date: 2013

Jay Papers Digital Collection

Sneak Preview at: <https://dlc.library.columbia.edu/jay>

THE PAPERS OF JOHN JAY

Start a new search...



ABOUT THE PAPERS OF JOHN JAY

The Papers of John Jay is an image database and indexing tool comprising some 13,000 documents (more than 30,000 page images) scanned chiefly from photocopies of original documents. Most of the source material was assembled by Columbia University's John Jay publication project staff during the 1960s and 1970s under the direction of the late Professor Richard B. Morris. These photocopies were originally intended to be used as source texts for documents to be included in a planned four-volume letterpress series entitled *The Selected Unpublished Papers of John Jay*, of which only two volumes were published.

BROWSE THE COLLECTION

LIBRARY LOCATION



NAME



YEAR



Ford IFP Archive

Sneak Preview at: <https://dlc.library.columbia.edu/ifp>



FORD FOUNDATION INTERNATIONAL
FELLOWSHIPS PROGRAM ARCHIVE

Start a new search...



IFP Countries ▾



ABOUT IFP

The IFP program began in November 2000 with four pilot sites. In Asia, the program was initiated in Vietnam. In Africa, the first

ABOUT THE COLLECTION

The archives cover the issues of social justice, community development, and access to higher education, and include paper and

IFP COUNTRIES



Brazil



Chile



China



Egypt

Additional Partnerships & Services

- ICPSR - <http://clio.columbia.edu/catalog/1360887>
- arXiv.org - <http://arxiv.org/>
- Social Sciences Data: <http://library.columbia.edu/locations/dssc/data.html>
- HathiTrust: <http://www.hathitrust.org/>
- Internet Archive: <https://archive.org/details/ColumbiaUniversityLibraries>



The Five Opens

- **Open Code** – sustainable platforms
- **Open Data Standards** – manage over time
- **Open API/Protocols** – foster interoperability
- **Open Communities** – go further together
- **Open License** – share knowledge broadly

Based on Mara Hancock's Three Opens



Questions

