Interoperability Between Institutional and Data Repositories: a Pilot Project at MIT

Katherine McNeill
Data Services and Economics Librarian
Massachusetts Institute of Technology
Introduction

- Social Science Data Services at MIT
- Support for MIT faculty as data producers
- Archiving and disseminating data
Multiple Locations for Depositing Data

- DSpace
- Harvard-MIT Data Center (HMDC): Runs on Dataverse Network Software (formerly Virtual Data Center (VDC))
- ICPSR
Presents Challenges For:

• Searching
• Unifying collections
• Archiving

• Therefore: Need for Interoperability
• Opportunity: PLEDGE Project
DSpace-Dataverse Interoperability

- Test case: MIT faculty-produced data in Dataverse from Jameel Poverty Action Lab (J-PAL)
- Goal: to archive, preserve, and provide access in DSpace to MIT-authored studies in HMDC Dataverse so that:
  - DSpace can archive the studies
  - Users can find the studies from within DSpace
  - Maintain access to unique services in Dataverse
- Developer: Mark Diggory, Systems Manager (former VDC developer at HMDC)
Technical Aspects of Agent

- Harvesting and replication of metadata and content from Dataverse into DSpace
- Converting DDI metadata into a form that can be understood and processed by DSpace
- METS (Metadata Encoding and Transmission Standard) SIP (Submission Information Package)
  - MODS descriptive metadata
  - PREMIS technical metadata
- Creation of an item in DSpace with DDI and study content as associated files
How it Works

Person sends URL

Study URL

DDI Record/Content

(Via OAI)

Agent

SIP

Ingest packager

DDSpace

Dataverse

Study URL
Demonstration: Item in Dataverse

Udaipur Health Study

Cataloging Information

How to Cite
Abhijit Banerjee; Angus Deaton; Esther Duflo, 2006, 'Udaipur Health Study', hdl:1902.1/RKYSWXPPAC UNF:3:skK7cR2Y+qClEyxD12Qg== Murray Research Archive [Distributor]

Study Global Id
hdl:1902.1/RKYSWXPPAC

Authors
Abhijit Banerjee; Angus Deaton; Esther Duflo

Producer
Abhijit Banerjee; Angus Deaton; Esther Duflo

Production Date
2006

Production Place
Cambridge, Massachusetts

Software
Stata

Funding Agency
Center for Health and Well-Being (Princeton University) and The John T. and Catherine MacArthur Foundation

Grant Number
SA3140, Center for Health and Well-Being (Princeton University) and The John T. and Catherine Mac Arthur Foundation
Demonstration: Item in DSpace

Title: Udaipur Health Study

Authors: Abhijit Banerjee
Angus Deaton
Esther Duflo

Keywords: Health Care, Health Care Facilities, Udaipur, Rajasthan

Issue Date: 26-Sep-2007

Abstract: This data set contains data on the health histories of, and access to healthcare facilities for
individuals in the Udaipur districts of Rajasthan, India. Data was collected at the household level, as well as
for adults and children. Also, private and public healthcare facilities were also surveyed.

URI: http://hdl.handle.net/123456789/10

Other Identifiers: hdl:1902.1/RKYSWXPPAC

Appears in Collections: J-PAL Datasets

Files in This Item:

<table>
<thead>
<tr>
<th>File</th>
<th>Description</th>
<th>Size</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFSData.dta</td>
<td></td>
<td>10 MB</td>
<td>Unknown</td>
</tr>
<tr>
<td>PFSData.pdf</td>
<td></td>
<td>315.02 kb</td>
<td>Adobe PDF</td>
</tr>
<tr>
<td>Questionnaires.zio</td>
<td></td>
<td>305.24 kb</td>
<td>Unknown</td>
</tr>
<tr>
<td>study.xml</td>
<td></td>
<td>11.42 MB</td>
<td>XML</td>
</tr>
</tbody>
</table>
Challenges

- Workflow for selecting and processing studies:
  - manual process
  - documentation
  - informal service
- Updating of studies
- License agreements and terms of use
- Keeping the agent up-to-date
Broader Implications

• Program can be adapted for other systems
• Interoperability of repositories based on different metadata conventions
• Packaging standards
• Management of complex digital objects
Possible Future Work

• Formalizing service
• Automation
• Bringing studies from DSpace into HMDC Dataverse
  – MIT faculty workflow
  – Dataverse features
  – Metadata complexity
Conclusion

• Prototype system enhances discovery and preservation of these data files
• Demonstrates how to package and import DDI metadata and data files into a greater variety of systems
• Lessons for importing complex digital objects into repositories