Dataverse Usability Evaluation: Findings & Recommendations

Presented by
Eric Gibbs
Lin Lin
Elizabeth Quigley
Agenda

- Introduction
- Scenarios
- Findings & Recommendations
- Next Steps
Introduction
Scenarios
Scenario One

- Data Users: Individuals who are exclusively searching for specific datasets, studies, collections, dataverses. Individuals who access and use datasets provided by other people.
## Scenario One

<table>
<thead>
<tr>
<th>Dataverse Feature</th>
<th>Task Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Searching (Basic and Advanced) and Refining</td>
<td>1, 2, 3a, 3b, 5a</td>
</tr>
<tr>
<td>Browsing dataverses</td>
<td>4a, 4b</td>
</tr>
<tr>
<td>Downloading Files (single &amp; multiple), Subsetting, Restricted Access</td>
<td>5b, 6a, 6c, 7</td>
</tr>
<tr>
<td>Citing Data</td>
<td>6b</td>
</tr>
<tr>
<td>User Guide</td>
<td>8</td>
</tr>
<tr>
<td>Icon Use</td>
<td>9</td>
</tr>
</tbody>
</table>
Scenario Two

- Data Creators: Individuals who upload, store, share and publish their own datasets, studies, collections
## Scenario Two

<table>
<thead>
<tr>
<th>Level</th>
<th>Dataverse Feature</th>
<th>Task Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account</td>
<td>Dataverse Creation</td>
<td>1</td>
</tr>
<tr>
<td>Dataverse</td>
<td>Settings- Editing Description</td>
<td>2</td>
</tr>
<tr>
<td>Account &amp; Dataverse</td>
<td>Studies – Study Creation</td>
<td>3</td>
</tr>
<tr>
<td>Study</td>
<td>Entering Cataloging Info</td>
<td>4</td>
</tr>
<tr>
<td>Account &amp; Study</td>
<td>Studies – Uploading Files</td>
<td>5</td>
</tr>
<tr>
<td>Study</td>
<td>Permissions – Restricting File Access</td>
<td>6</td>
</tr>
<tr>
<td>Study</td>
<td>Release study</td>
<td>7</td>
</tr>
<tr>
<td>Dataverse</td>
<td>Collection – Collection Creation (Static &amp; Dynamic)</td>
<td>8a, 8b</td>
</tr>
<tr>
<td>Dataverse</td>
<td>Settings – Publishing Dataverse</td>
<td>9</td>
</tr>
<tr>
<td>Homepage &amp; Study</td>
<td>Re-access Dataverse</td>
<td>10</td>
</tr>
<tr>
<td>Dataverse</td>
<td>Settings- Unrelease Dataverse</td>
<td>10</td>
</tr>
</tbody>
</table>
Findings & Recommendations
Pre and Post Test Survey

- Twenty-one users were involved in testing, two of them were pilot participants

- 12/19, 5/19, 2/19

- 4.2/7

- 13/19
Basic & Advanced Search

- Users understood where to locate the basic search function well

- Advanced search was more problematic
  - Users were frustrated that when they received no search results they could not go back and edit what they already had
  - They were not aware of what would happen when they hit enter on the advanced search page
  - The advanced search fields were hard for users to grasp
Browsing dataverses

**Dataverses**

A *Dataverse* is a container for research data studies, customized and managed by its owner.

---

**BROWSE DATAREVRESSES BY FILTER BY ORGANIZATION**

- Association, Society: 37
- Government: 24
- School (Other): 194
- Business, Industry: 5
- School (Harvard): 101

---

**BROWSE DATAREVRESSES BY FILTER BY TYPE**

- Institutions, Depts. Centers: 74
- Large Collections: 33
- Scholars: 265
- Journal Replication Archive: 30
- Research Projects: 114

---

**RECENTLY RELEASED DATAREVERSSES**

- open: May 10, 2013
- Heppler, Michael: May 3, 2013
- gdp: Apr 23, 2013
- es_test_34: Apr 19, 2013

View More >
Accessibility

- Participants commented on font size being too small
- Red and green asterisks denoting required vs. recommended fields on forms hard to either see or differentiate

Recommendation:

- Work towards creating a version of Dataverse that follows accessibility guidelines
- Harvard’s University Disability Services recommends following the W3C’s Web Accessibility Initiative Guidelines
Identifying specific studies & data files

- Data & Analysis Tab
  - Font size issues
  - Difficulty distinguishing between files
  - Loading more files on the page

- Downloading subsets of variables
  - Ability to sort variables
  - Downloading options
  - Citing a data file
Identifying specific studies & data files

Use the check boxes next to the file name to download multiple files. Data files will be downloaded in their default format. You can also download all the files in a category by checking the box next to the category name. You will be prompted to save a single archive file. Study files that have restricted access will not be downloaded.

Due to the large number of files associated with this study, only 25 files are loaded at a time.

<table>
<thead>
<tr>
<th>Data and Documentation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SAP 1996-2000 Pupil Questionnaire, 1997-1999 Pupil Interview.zip</strong></td>
<td>Files in original formats, zipped file</td>
</tr>
<tr>
<td><strong>SAP Class Observation, Teacher Questionnaire, 1996-1999.zip</strong></td>
<td>Files in original formats, zipped file</td>
</tr>
<tr>
<td><strong>SAP School Committee [1996-2000], Textbooks [1999].zip</strong></td>
<td>Files in original formats, zipped file</td>
</tr>
<tr>
<td><strong>School Assistance Program in Rural Kenya 1996-2000 Part 1 1996-1998.zip</strong></td>
<td>Files in original formats, zipped file</td>
</tr>
<tr>
<td><strong>School Assistance Program in Rural Kenya 1996-2000 Part 2 1999-2000.zip</strong></td>
<td>Files in original formats, zipped file</td>
</tr>
<tr>
<td><strong>School Assistance Program in Rural Kenya Test Scores I 1996-2000.zip</strong></td>
<td>Files in original formats, zipped file</td>
</tr>
<tr>
<td><strong>School Assistance Program in Rural Kenya Test Scores II 1996-2000.zip</strong></td>
<td>Files in original formats, zipped file</td>
</tr>
</tbody>
</table>

**2a. SAP 1996-2000 Pupil Questionnaire, 1997 - 1999 Pupil Interview: Data Files**

<table>
<thead>
<tr>
<th>Data File</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>98pupil1_2.tab</strong></td>
<td>Tab Delimited - 4 MB - 131 downloads + analyses</td>
</tr>
<tr>
<td><strong>98pupil2_2.tab</strong></td>
<td>Tab Delimited - 3 MB - 74 downloads + analyses</td>
</tr>
</tbody>
</table>
## Identifying specific studies & data files

<table>
<thead>
<tr>
<th>Study/Document</th>
<th>Data Type</th>
<th>Cases</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>YU_Pupint3.tab</td>
<td>Tab Delimited</td>
<td>2531</td>
<td>76</td>
</tr>
<tr>
<td>98PupQ1.tab</td>
<td>Tab Delimited</td>
<td>2530</td>
<td>131</td>
</tr>
</tbody>
</table>

- Pupint97.doc  
  - MS Word - 18 KB - 76 downloads
- Pupint98.doc  
  - MS Word - 40 KB - 68 downloads
- QU_PUP99.DOC  
  - Plain Text - 43 KB - 70 downloads

- 98pupill1_2.pdf  
  - Adobe PDF - 384 KB - 79 downloads
- 98pupill2_2.pdf  
  - Adobe PDF - 198 KB - 66 downloads
- 97pupint_2.pdf   
  - Adobe PDF - 135 KB - 46 downloads
- 97pup_T101.pdf   
  - Adobe PDF - 394 KB - 70 downloads

[Load Next 25 Files] [Show All Files]
## Identifying specific studies & data files

### Variable Information Table

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
<th>Label</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous</td>
<td>schoolid</td>
<td>ID Number of school</td>
<td></td>
</tr>
<tr>
<td>Discrete</td>
<td>studlid</td>
<td>ID Number of student</td>
<td></td>
</tr>
<tr>
<td>Continuous</td>
<td>std</td>
<td>Standard student is in</td>
<td></td>
</tr>
<tr>
<td>Discrete</td>
<td>day</td>
<td>Day of the interview</td>
<td></td>
</tr>
<tr>
<td>Discrete</td>
<td>month</td>
<td>Month of the interview</td>
<td></td>
</tr>
<tr>
<td>Discrete</td>
<td>year</td>
<td>Year of the interview</td>
<td></td>
</tr>
<tr>
<td>Character</td>
<td>language</td>
<td>Mother tongue</td>
<td></td>
</tr>
<tr>
<td>Character</td>
<td>religion</td>
<td>Religion</td>
<td></td>
</tr>
<tr>
<td>Character</td>
<td>religoth</td>
<td>Religion if 'other' option mark</td>
<td></td>
</tr>
<tr>
<td>Character</td>
<td>nursery</td>
<td>Nursery class before std1</td>
<td></td>
</tr>
</tbody>
</table>
Creating an account & dataverse

- Multiple participants expected clicking on the create an account box would link them directly to the create account page.

- Error messages would sometimes occur off screen and participants would not know where they had made an error.
Creating studies & uploading files

- Participants noted the asterisks on the create study form as being difficult to differentiate

- The instructions at the top of the page blended into the rest of the content

- Using “upload” instead of “save” when uploading files
Settings & Permissions

- Restricting Access for Files
  - Participants weren’t sure which section accurately restricted files
  - Participants weren’t confident in the terminology used for restricting access
## Settings & Permissions

**MANAGE STUDY PERMISSIONS**

Restrict or permit access to this study, all files uploaded to the study, or selected files only. You can restrict all users, or permit selected users to access the study or files.

### ENTIRE STUDY PERMISSION SETTINGS

Cataloging Information and File Settings
- **Public**: If this setting is public, files can still be restricted below.

User Restricted Study Settings
- **Username/Group**: Enter username/group to allow restricted study access.

### FILE PERMISSION SETTINGS

Restricted File Settings
- **Allow users to request access to restricted files for this study** (you will get an email notification for each request).

### INDIVIDUAL FILE PERMISSION SETTINGS

Restricted File User Access
- **Select file and enter username/group to allow restricted file access, or set the file permission**. (The single update button works for either field.) Select multiple users/groups by checking the boxes under User/Group Restricted File Access.

<table>
<thead>
<tr>
<th>Category Name</th>
<th>File Name</th>
<th>Found In Version(s)</th>
<th>Permission</th>
<th>User/Group Restricted File Access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Save**  **Cancel**
Settings & Permissions

- Releasing a dataverse or study
  - Participants weren’t confident they had actually released their dataverse
  - Participant stated it would be nice to have a confirmation pop up appear to confirm releasing of study

- Streamline Permissions
  - Condense permissions pages for studies and dataverses into one page
  - Users will no longer click around looking for the separate permissions pages
Collections

- Participants had difficulty locating the Collections tab
  - User guide was not helpful as path to tab was not included
- Participants had easier time creating dynamic collection after static collection creation
  - Familiar with process and had located Collections tab
- “Child” and “parent” terminology caused confusion
  - Participants were not familiar with terminology
- Collections tab is labeled “Sub-Collection” once it is clicked on
  - Caused participants to think they were in the wrong tab
Collections

- Static collection creation
- Participants didn’t notice list of studies at bottom of page
Homepage

- Dataverse home link is hard to identify
  - Users often tried the “IQSS” link in the top left corner first, followed by the “Powered by the Dataverse Network” on the right side
  - Users often did not identify the Dataverse homepage link as a link
User Guides

- Users liked text in the guides as opposed to screen shots or videos

- The guide content does not currently match Dataverse’s content in all cases

- Users did not like that the guide pops up in a different window
Icons

- Tested with both scenarios

- Search and user guide icons easy to identify

- Report a problem icon not very intuitive
  - Participants thought it was a cartoon bubble or chat icon

- Plain text rather than icons
  - “It would be more straightforward than the icons”-participant
Icons

- Settings icon
  - Does not appear on Dataverse homepage once a user is logged in
  - Takes 3 clicks before it appears

- Recommendation:
  - Persistent settings icon
Icons
Icons
Icons
Hierarchy of Dataverse

- Users had difficulty visualizing and explaining hierarchy

- Confused by how collections fit in Dataverse

- Recommendation: add a visual representation of Dataverse
  - Content uploaders can easily tell where they are in Dataverse
  - Content uploaders can easily access different levels of their content
Accounts

- Differences between administrator account and basic account
Accounts

Harvard Dataverse Network

The Harvard Dataverse Network is open to all scientific data worldwide. Includes the world's largest collection of social science research data. Learn more about the Dataverse Network.

Dataverses

528 Dataverses

- A Dataverse is a container for research data studies, customized and managed by its owner.

BROWSE DATATURES BY FILTER BY ORGANIZATION

- Association, Society (37)
- Business, Industry (5)
- Government (24)
- School (Harvard) (101)
- School (Other) (194)

BROWSE DATATURES BY FILTER BY TYPE

Studies

51,816 Studies, 720,908 Files, 770,557 Downloads

- A study is a container for a research data set. It includes cataloging information, data files and complementary files.

RECENTLY RELEASED STUDIES

- open_st
  - May 16, 2013
- wiki
  - May 10, 2013
- Notification Test by Happier, Michael
  - May 3, 2013

Share, Cite, Reuse, Archive Research Data
Scientific data for reproducible research

Powered by the Dataverse Network v. 3.4

Usability Testing Log Out
Next Steps
Further Analysis of User Testing Sessions

- Our analysis is based primarily on qualitative data

- A thorough analysis of time on tasks, user difficulties, and user failures would flesh out our observations
Analytic Data Evaluation

- Dataverse currently keeps analytic data on site usage

- Analytics would show what browsers are most commonly used, how long users spend on pages, and how often the account creation page is accessed

- This information will help guide developers in the future
Persona Development

• What is a persona?
  • Includes: age, fake name, hobbies, work environment, etc
  • Helps to understand how users are interacting with Dataverse

• Data creator and data user current user profiles
  • Need to be expanded, i.e.-staff, faculty, students? Who are they?

• Usability method: survey
Faceted Search Competitive Analysis

- Faceted search for next iteration of Dataverse would benefit from a competitive analysis

- Competitive analysis of databases featuring faceted search used by Dataverse users
  - How is their faceted search set up?
  - What filters do they use?
Dataverse & Research Site Linking

- Users commonly use Google Scholar or HOLLIS to locate studies with data sets contained in Dataverse

- Linking Dataverse studies to their finished products in commonly used search engines would help the site’s traffic
Further User Testing

- Dataverse creators were underrepresented in this user testing run

- Remote testing would allow Dataverse to solicit more information from data creators