Research Data Management Best Practices and REDCap

Paul A. Harris, PhD, FACMI
Professor, Biomedical Informatics
Director, Office of Research Informatics

Hosted by
Librarians are increasingly involved in guiding researchers in data management, given that it has become a requisite for some granting institutions and publications to provide a data management plan and provide data together with the articles.

The following are some guidelines to data management that will be useful for librarians to know in advising, assisting or cooperating with researchers on data management and an example of a not-for-profit data management tool.
Important Concept

Data planning is a wise investment for any research study ...

Garbage In...

... Garbage Out
Always, clearly define all data requirements before seeing the first research subject.

It’s an ethics thing!
Good Practice

Develop a primary hypothesis, then fully define primary and secondary outcome measurements necessary to test hypothesis.
Good Practice

Be exact and discrete when defining variables.

This is a really simple study ...

---- HINT: think about a representative number + units ...

Blood Pressure → Systolic BP (mmHg) + Diastolic BP (mmHg)
Metabolic Panel → Individual Measurements (with units)
Medications → Beta-Blocker (Yes/No) and/or Dosage
Good Practice

Code categorical variables during data collection. It helps in later analysis.

**Gender:** 0 = Male, 1 = Female

**Education:**
- 0 = no high school;
- 1 = high school;
- 2 = some college;
- 3 = college degree

Be consistent! Similar questions should use similar codes.

Keep a current copy of the codebook for use during statistical analyses.
Good Practice

Always include a set of confounding factors (e.g. race, gender, date of birth → age, height, initial weight, smoker, etc) when planning data collection.

These data may be used later to adjust for factors in final data analysis or to check validity of randomization.

Consider DNA collection if appropriate for the study. Collection is cheap (you don’t have to sequence) and may later help identify study outliers even if you do not initially suspect a genetic component.

* Remember human subjects board + local policy requirements.
Allow for the possibility of missing data in data fields. Never code a zero or other possible value as default.

Develop codes for missing values if context is needed

  ... Patient doesn’t remember (888)
  ... Patient will not disclose (999)

Dates can be particularly problematic...
When were you diagnosed?

  April 1 2003; Spring 2003; 2003; In last 5-years...
Data Sharing / Exchange

Sharing data requires a well-defined data dictionary (including standards encoding where possible) and standard operating procedures.

Data coding standards are not well-defined in every scientific research domain, but you can always maintain a structured data dictionary for a study.
Data Integrity and Validation

Use software that checks data type and ranges when accepting data from end-user.

Double-data entry (two pass verification) can be useful when accuracy is especially important.
Good Practice

Keep audit trails of all data changes.

Keep logs of who added or changed data.

Same principles apply for paper and electronic records ...
Want more?

https://www.coursera.org/learn/clinical-data-management/

- Covers basic concepts of designing a good data collection strategy
- Self-paced (no hard deadlines)
- Free!
Why REDCap?
The Software Platform
2004 Gap Assessment

Researchers were using sub-standard methods to manage data for research studies/trials projects (pilot, R01, PPG)
Visual Status

Data Validation

Numerous Field Types

Human Readable Labels

Auto-Variable Coding

Branching Logic

PDFs

+ Text (Free)
  + Number
  + Phone
  + Zip
  + Date
  + TextArea
  + Select
  + Radio
  + File
IMPLEMENTATION EXERCISE

https://projectredcap.org/software/try/
POWER FROM THE PEOPLE

A STORY

Power to the People

Paul Harris – REDCapCon 2016
Why REDCap?

REDCap filled a critical gap at Vanderbilt

Easy way to do the right thing ...
Early QI Survey (N=20) – Did REDCap improve your science?

Yes (N=20)

- Improving Data
- Improves Science
- Improves Health
Empowered users creating projects solidifies the data management plan before enrolling the first subject.

Improving Data

Improve Science

Improves Health

2004 Survey --- N = 20 Early Adopters REDCap Improves Science! (100%)
Is a unique, supportive community and the power does flow in ALL directions

But here’s the cool POWER part ...
So ... how's it working so far?

It's fantastic, but could you add some sort of widget to ...
POWER

Used systems get better ...

*End-users will tell you what you need to know to improve if you listen*
Local @ VU

Typical Week

Available to any student, faculty, staff member for any Vanderbilt use – no cost

System Statistics

<table>
<thead>
<tr>
<th>Total projects (excludes 'practice' projects)</th>
<th>40,551</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Project status</td>
<td></td>
</tr>
<tr>
<td>- Production</td>
<td>15,527</td>
</tr>
<tr>
<td>- Development</td>
<td>15,186</td>
</tr>
<tr>
<td>- Inactive</td>
<td>2,285</td>
</tr>
<tr>
<td>- Archived</td>
<td>7,553</td>
</tr>
<tr>
<td>• Project types</td>
<td></td>
</tr>
<tr>
<td>- Projects not containing surveys</td>
<td>10,961</td>
</tr>
<tr>
<td>- Projects containing surveys</td>
<td>29,590</td>
</tr>
<tr>
<td>• Project purpose</td>
<td></td>
</tr>
<tr>
<td>- Research</td>
<td>11,928</td>
</tr>
<tr>
<td>- Basic or bench research</td>
<td>1,629</td>
</tr>
<tr>
<td>- Clinical research study or trial</td>
<td>4,425</td>
</tr>
<tr>
<td>- Translational research 1 (applying discoveries to the development of)</td>
<td>603</td>
</tr>
<tr>
<td>- Translational research 2 (enhancing adoption of research findings at)</td>
<td>481</td>
</tr>
<tr>
<td>- Behavioral or psychosocial research study</td>
<td>2,911</td>
</tr>
<tr>
<td>- Epidemiology</td>
<td>748</td>
</tr>
<tr>
<td>- Repository (developing a data or specimen repository for future use)</td>
<td>574</td>
</tr>
<tr>
<td>- Other</td>
<td>1,122</td>
</tr>
<tr>
<td>- [Subcategory not selected]</td>
<td>926</td>
</tr>
<tr>
<td>- Operational Support</td>
<td>10,880</td>
</tr>
<tr>
<td>- Quality Improvement</td>
<td>12,718</td>
</tr>
<tr>
<td>- Other</td>
<td>4,581</td>
</tr>
</tbody>
</table>

Project attributes (all projects)

- Data collection instruments: 124,248
- Fields on all data collection instruments: 5,404,250
- Records: fetch
- Survey responses: 2,030,525
- Survey participants (in participants list): 989,203
- Active users: 19,672
and let me know if you’re interested in collaborating ...
Consortium Partners

2,127 institutions in 108 countries
309,000 projects with 434,000 users

Map of REDCap Consortium Partners

Available at no cost to academic, non-profit, government
Long ago ... before we had real professionals doing this work ...

So ... how's it working so far?

It's fantastic, but could you add some sort of widget to ...
Power to the People
List of 2842 publications that have cited REDCap:

Displaying first 20 citations of 2842.
Quantifying biliverdin in liver and spleen samples from multiple avian species
Butler MW, Bociulis SC, Little AR, Minnick JA, Ritter NJ, Rockman ME, Rossi ML, Hines JK.

The US Initiative: Clinical and Translational Science Awards-The UC Davis Perspective
Tarantal AF, Rainwater J, Wun T, Berglund L. Book editors: Alving B, Dai K, Chan SHH.

Novel Emergency Department Risk Score Discriminates Acute Coronary Syndrome Among Chest Pain Patients With Known Coronary Artery Disease

Prevalence and factors associated with the absence of pharmacologic venous thromboembolism prophylaxis: A cross-sectional study of Georgia intensive care units
Hawkins A, Mazzeffi M, Abraham P, Paciullo C.

Clinical Study of an Online Tool for Standardizing Hospital Care
Vanderbilt Information Scientist

Taneya Y Koonce, MSLS, MPH
Phone: (615) 936-5790
Email: taneya.koonce@Vanderbilt.Edu
Office: 223 Eskin Biomedical Library

Job Related Information

Job Title:
- Senior Information Scientist

Department:
Center for Knowledge Management

Area(s) of Responsibility:
- Administrative Services
- Research Services
- Education Services
- Knowledge Management
- Filtering/Evidence-Based Services

Statistics & Other Information

User currently logged in? No
Projects user can access (view projects) 73
Time of account creation 10/05/2009 4:17pm
Time of first login 10/05/2009 4:17pm
Time of last login 12/01/2016 4:44pm
Time of first activity 10/05/2009 4:18pm
Time of last activity 11/30/2016 3:59pm
The Center for Knowledge Management at Vanderbilt University Medical Center recently conducted the first-known feasibility study of the Institute of Medicine-recommended panel of social and behavioral determinants of health. Using a randomized, parallel design, we tested the effects of question order in an online ResearchMatch sample of close to 800 volunteers from across the country. Our findings confirmed the feasibility and reliability of the IOM SBD panel of questions, thus providing reassurance for the questionnaire’s use by providers to systematically capture patient social and behavioral health needs. Our team leveraged REDCap to administer the questionnaire (three different versions sent to participants approximately one week apart), and to serve as the repository for all collected data. (Giuse NB, Koonce TY, Kusnoor SV, Prather AA, Gottlieb LM, Huang LC, Phillips SE, Shyr Y, Adler NE, Stead WW. Institute of Medicine Measures of Social and Behavioral Determinants of Health: A Feasibility Study. Am J Prev Med (2016), http://dx.doi.org/10.1016/j.amepre.2016.07.033).
THE LIBRARIAN’S ROLE IN TRANSLATIONAL RESEARCH

Hardcover | August 29, 2016

Editor Marisa L. Conte

★★★★★ not yet rated | write a review

In 2005, the National Institutes of Health (NIH) announced the Clinical and Translational Science Award (CTSA) program - an institution-based grant intended to re-engineer the clinical research enterprise, speeding the time from pre-clinical discovery to the development of therapies to improve human health. As universities competed for CTSA funding and often struggled to develop or recalibrate institutional infrastructures and research support services, the face of pre-clinical and clinical research changed dramatically. These

TITLE: Librarian Integration in a Working Group of the REDCap International Consortium

Authors: Jennifer A. Lyon¹, MS, MLIS, Fatima M. Mncube-Barnes², Ed.D., M.P.H., MSIS, Brenda L. Minor³, BS

¹Health Sciences Library, Health Sciences Center, Level 3, Stony Brook University, Stony Brook, NY 11794-8034 jennifer.a.lyon@stonybrook.edu

²Meharry Medical Library, Stanley S. Kresge Learning Resources Center, 1005 Dr. D.B. Todd, Jr Blvd. Nashville, TN 37208 barnes@mmc.edu

³The Vanderbilt Institute for Clinical and Translational Research, 2525 West End Ave, 6th Floor, Nashville, TN 37203-8820 brenda.l.minor@vanderbilt.edu

Abstract

With the rapid growth of translational science, novel opportunities have arisen for biomedical librarians to contribute in unique ways. Two librarians who actively supported research initiatives at their institutions using REDCap (Research Electronic Data Capture), a free globally-utilized software system for collecting and storing research data, were invited to assist in the development of an electronic data collection instrument library application (the Shared Data Instrument Library) housed in REDCap. The librarians assisted with the instrument curation process, governed by the REDCap Library Oversight Committee (REDCOC), by assuming primary investigative roles for determining copyright ownership and subsequently obtaining permission from copyright holders to adapt instruments into REDCap. Over time, the

When the librarians came on board with the REDLOC Committee in December 2010 and January 2011, respectively, the SDIL contained 80 instruments with a total of 383 downloads. As of June 8, 2015 there were 907 instruments (655 curated + 252 user submitted) and over 15,000 downloads.
Any Questions?