Apr 9th, 1:45 PM

Tales from a Data Management Survivalist: Skills Honed in the Wilderness

Karen L. Hanson
New York University School of Medicine

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium

Part of the Scholarly Communication Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License.

Hanson, Karen L., "Tales from a Data Management Survivalist: Skills Honed in the Wilderness" (2014). University of Massachusetts and New England Area Librarian e-Science Symposium. 4.
https://escholarship.umassmed.edu/escience_symposium/2014/program/4

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
Tales from a data management survivalist: Skills honed in the wilderness

New England e-Science Symposium
April 9, 2014

Karen Hanson
Knowledge Systems Librarian
karen.hanson@med.nyu.edu
Sorry

(I’m a medical librarian)
Something that inspires and scares me

“Don’t assume that people care about libraries. People care about streamlining the processes that support research and learning.”

http://www.ala.org/acrl/issues/value/changingroles
Data services: where to start?
Naked and afraid in the data wilderness
Library’s data strengths (2011)

- Resources: 2
- Knowledge: 3
- Stamina: 8
• Naked and Afraid
• **Dropped in the jungle**
• Honing our survival skills
• Paddling down the river
• Lessons learned
Environmental scan

- Complex environment
- Lots of small isolated services
- Lots of gaps / opportunities
A starting point: Education (Sept 2011)

- First step to building a résumé
- Learn about what people need
- Demonstrate our understanding
- Test the water!
Creating an opportunity

- Contacted postdoctoral program director
- 90 minute class:
  - Plant seeds of thought
  - Raise awareness
  - Give practical pointers for immediate improvements
Class outline

• Introduction
• Incentives (carrots & sticks)
• Standards for description & documentation
• Storage, archiving and sharing
• Data management planning
Class features: Scare tactics

Government mandates timeline

2003: NIH adopted a data sharing policy.

2008: NIH implements the Public Access Policy

(still no teeth, but young yet)

Government mandates timeline

2011: NSF made data management plans a requirement

2013: NIH Public Access Policy... now with teeth
“There were 60 children in the study. Ages were by accident duplicated between the upper and lower halves of the database. Thus, the ages for the first 30 children in the data set were identical and in the same order with the ages for the second set of 30 children... The files with the original data are not available any more, making it impossible to reconstruct a valid data set for reanalysis.”

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3320558/
Class features: Real world examples

1 rat heart
100s of slices
100s of slides
1000s of image files
100s of huge images

5-7 experiments a week...

3 post docs
Class features: Postdoc survey

- ~2500 responses from 43 institutions analyzed
- 3 questions on data management

**To what extent have you dealt with NIH data sharing regulations or NSF data management plans?**

<table>
<thead>
<tr>
<th>Not aware of policies</th>
<th>Aware but no involvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>38%</td>
<td>48%</td>
</tr>
<tr>
<td>39%</td>
<td>48%</td>
</tr>
</tbody>
</table>

**NYU Health Sciences Library**

Introduction to Data Management

Pre-class Survey

1. Data management and sharing policies from federal funding agencies, such as NIH and NSF, have been implemented in order to ensure that research data is accessible to other investigators. To what extent have you dealt with NIH data sharing regulations or NSF data management plans? Check all that apply.

- [ ] I have had to write a data sharing/management plan for a grant proposal
- [ ] I have had to implement data sharing/management plan in my lab
- [ ] I am aware of federal funding agency policies around data sharing/management, but have not had any involvement with them
- [ ] I am not aware of federal funding agency policies around data sharing/management

http://hsl.med.nyu.edu
Class features: Chilling tales from our own lives
Class features: Humor
Class evaluation

Will you use the topics covered in your work?

- Definitely will
- Probably will
- Probably won't
- Definitely won't

Would you be interested in future classes that went into more detail?

- Yes
- No
June 2012
• Naked and Afraid
• Dropped in the jungle
• **Honing our survival skills**
• Paddling down the river
• Lessons learned
Researcher experience of data support at our institution?
Division of Knowledge Informatics (DKI)
Funding announcement

NLM Administrative Supplements for Informationist Services in NIH-funded Research Projects
The grant

“Clinical Management of Cochlear Implant Patients with Contralateral Hearing Aids”
Mario Svirsky & Arlene Neuman
The informationist supplement

- Data model / database
- Data entry tool
- Refine reporting queries
- Query tool

Informationists:
- Theodora: data modeling
- Me: database programming, application design
Domain knowledge
Subjects

Research Team

The Data

MS Access Database

Principal Investigators
The Data

Subjects
Research Team

Principal Investigators
October 2012: Hurricane Sandy
After
Taking one on the chin

- Knowledge
- Resources
- Stamina

Bar chart showing Stamina with the highest value, followed by Knowledge and Resources.
Naked and afraid
A glimmer of hope
• Naked and Afraid
• Dropped in the jungle
• Honing our survival skills
• Paddling down the river
• Lessons learned
A fork in the river

basic

clinical
Basic to clinical: Apples to oranges

**Basic scientists:**
- Much wider variety of data
- Data practices… the wild west
- Postdocs

**Clinical investigators:**
- Data more consistent
- Systems available (e.g. REDCap, Velos)
- Greater recognition of value in sharing
Basic scientists - strategy

1) Continue integration into postdoc programs
Basic scientists - strategy

2) Keep improving existing material

Metadata – general structure

- Dublin Core
  - Designed to be generic/flexible
  - Usually stored as XML
    - e.g. `<doc creator>Harison, Karen L.</doc>
  - 15 fields:
    - Contributor, Coverage, Creator, Date, Description, Format, Identifier, Language, Rights, Source, Title, Type, Version, Rights Holder, RelatedURL
Basic scientists - strategy

3) Seek out new opportunities through liaisons
Clinical investigators – strategy

1) Partner with existing expert
Clinical investigators – strategy

2) Create short modules for busy clinicians

- Module #0 - How to avoid a data management nightmare (teaser)
- Module #1 - Introduction to Data Management
- Module #2 - Planning Data Collection
- Module #3 - Data Structure and Naming Conventions
- Module #4 - Form Design
- Module #5 - Electronic Data Capture
- Module #6 - Data integrity monitoring
- Module #7 - Analysis
- Module #8 - Privacy issues
- Module #9 – FDA / FISMA
- Module #10 – How to document your data (and why!)
- Module #11 – Storage, Preservation
- Module #12 – Sharing
Clinical investigators – strategy

3) Participate in new workgroup to develop education program for clinical investigators

How to avoid a data management nightmare

NYU Health Sciences Library
Karen Hanson | Kevin Read | Alisa Surkis
Meanwhile, the informationist project
Subjects Research Team Principal Investigators

International Researchers

The Data

New Database

MS Excel
Tool evaluation
Will we ever get this thing started?
Original data entry tool

![Data entry form screenshot]

**DAT_Suject Form**

<table>
<thead>
<tr>
<th>Tab</th>
<th>Data</th>
</tr>
</thead>
</table>
| Subject Data| **Subject ID:** 5
|             | **Last Name:**              |
|             | **FirstName:**              |
| Parent/Guardian|                            |
| Left Implant| **Implant ID:** 7
|             | **Ear:** R                   |
|             | **Implant Device:** Tempo 40|
|             | **Implant Date:** 1/1/2011   |
|             | **No. of Electrode:** 22     |
|             | **Stimulation Date:** 1/1/2011|
|             | **Active:**                  |
| Right Implant|                            |
| Hearing Aids|                            |
| Test Scores |                            |
| Other Info  |                            |

**Processor Detail**

<table>
<thead>
<tr>
<th>Processor ID</th>
<th>Processor</th>
<th>Strategy ID</th>
<th>Strategy</th>
<th>Strategy Date</th>
<th>Active</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Esprit</td>
<td>9</td>
<td>ACE</td>
<td>1/1/2011</td>
<td></td>
</tr>
</tbody>
</table>

**Strategy Detail**

- Processor ID: 7
- Processor: Esprit
- Strategy ID: 9
- Strategy: ACE
- Strategy Date: 1/1/2011
- Active: Yes

**Show - Hide Strategy**

**Show - Hide Processor**

**Subject Browse**

**Single Subject Search**
Tool evaluation
OK, we’re in it for the long haul
A unified model
Cleaner data entry
Validation, autocomplete, audit
### Built-in and Custom Reporting

#### Built-in Reports

<table>
<thead>
<tr>
<th>Report ID</th>
<th>Report Name</th>
<th>Last Run</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Intelligibility</td>
<td>09/30/2013</td>
</tr>
<tr>
<td>4</td>
<td>Expressive</td>
<td>12/01/2013</td>
</tr>
<tr>
<td>3</td>
<td>Bimodal</td>
<td>12/01/2013</td>
</tr>
<tr>
<td>2</td>
<td>General</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>CI report</td>
<td></td>
</tr>
</tbody>
</table>

#### Custom Reports

<table>
<thead>
<tr>
<th>Report ID</th>
<th>Report Name</th>
<th>Score</th>
<th>LeftInitImpDat</th>
<th>RightInitImpDat</th>
<th>TestDate</th>
<th>LeftTestDev</th>
<th>RightTestDev</th>
<th>Score</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>158</td>
<td>International subj</td>
<td></td>
<td>2012-01-0</td>
<td>2008-04-03</td>
<td>2009-03-0</td>
<td>No Aid</td>
<td>No Aid</td>
<td>75</td>
<td>AudioMetric 2000 UnAided Right</td>
</tr>
<tr>
<td>158</td>
<td>CI subgroup</td>
<td></td>
<td>2012-01-0</td>
<td>2008-04-03</td>
<td>2008-02-2</td>
<td>No Aid</td>
<td>No Aid</td>
<td>85</td>
<td>AudioMetric 4000 UnAided Left</td>
</tr>
<tr>
<td>158</td>
<td>Bimodal recent</td>
<td></td>
<td>2012-01-0</td>
<td>2008-04-03</td>
<td>2008-02-2</td>
<td>No Aid</td>
<td>No Aid</td>
<td>45</td>
<td>AudioMetric 250 UnAided Right</td>
</tr>
<tr>
<td>158</td>
<td>Potential subjects</td>
<td></td>
<td>2012-01-0</td>
<td>2008-04-03</td>
<td>2008-02-2</td>
<td>No Aid</td>
<td>No Aid</td>
<td>65</td>
<td>AudioMetric 1000 UnAided Left</td>
</tr>
<tr>
<td>158</td>
<td>Test report 1</td>
<td></td>
<td>2012-01-0</td>
<td>2008-04-03</td>
<td>2008-03-0</td>
<td>No Aid</td>
<td>No Aid</td>
<td>65</td>
<td>AudioMetric 1000 UnAided Right</td>
</tr>
<tr>
<td>158</td>
<td></td>
<td></td>
<td>2012-01-0</td>
<td>2008-04-03</td>
<td>2008-02-2</td>
<td>No Aid</td>
<td>No Aid</td>
<td>50</td>
<td>AudioMetric 250 UnAided Left Ear</td>
</tr>
<tr>
<td>158</td>
<td></td>
<td></td>
<td>2012-01-0</td>
<td>2008-04-03</td>
<td>2008-02-2</td>
<td>No Aid</td>
<td>No Aid</td>
<td>90</td>
<td>AudioMetric 2000 UnAided Left</td>
</tr>
<tr>
<td>158</td>
<td></td>
<td></td>
<td>2012-01-0</td>
<td>2008-04-03</td>
<td>2008-02-2</td>
<td>No Aid</td>
<td>No Aid</td>
<td>60</td>
<td>AudioMetric 1000 UnAided Left</td>
</tr>
<tr>
<td>158</td>
<td></td>
<td></td>
<td>2012-01-0</td>
<td>2008-04-03</td>
<td>2008-03-0</td>
<td>No Aid</td>
<td>No Aid</td>
<td>55</td>
<td>AudioMetric 250 UnAided Right</td>
</tr>
<tr>
<td>158</td>
<td></td>
<td></td>
<td>2012-01-0</td>
<td>2008-04-03</td>
<td>2008-02-2</td>
<td>No Aid</td>
<td>No Aid</td>
<td>50</td>
<td>AudioMetric 1000 UnAided Right</td>
</tr>
<tr>
<td>158</td>
<td></td>
<td></td>
<td>2012-01-0</td>
<td>2008-04-03</td>
<td>2008-03-0</td>
<td>No Aid</td>
<td>No Aid</td>
<td>90</td>
<td>AudioMetric 2000 UnAided Right</td>
</tr>
</tbody>
</table>
Informationist supplement – take aways

- Available tools
- Researcher workflows
- Contacts in Research IT
- Valuable, but select projects carefully
• Naked and Afraid
• Dropped in the jungle
• Working on our skills
• Paddling down the river
• Lessons learned
Post-evaluation of skills

- Knowledge
- Resources
- Stamina

The chart shows the evaluation of skills in terms of knowledge, resources, and stamina.
Challenges: Outside of our comfort zone
Challenges: Time, effort, persistence
We had no idea where to start

informationist

grant

education
Used library strengths

- Scholarly communication issues
- Repositories, data sharing
- Education
- Subject specialists / liaisons
- Metadata
- Finding answers
Used individual strengths
Forged partnerships

- Data needs are *enormous*!
- Partnerships make us stronger
- We can bring something to the table
Experienced pockets of success
To be continued…

You are here
Acknowledgements

NYU School of Medicine

Librarians:
- Theodora Bakker
- Kevin Read
- Alisa Surkis
- Neil Rambo

Researchers:
- Mario Svirsky
- Arlene Neuman

Grant supplement funders
- NLM
- NICDD
References


Hanson, K, Surkis, A, & Read, K. “Introduction to Data Management” http://hsiguides.med.nyu.edu/data_management

Hanson, K, Read, K, & Surkis, A, “How to avoid a data management nightmare” https://www.youtube.com/watch?v=nNBiCcBlwRA

Hanson, K, Surkis, A, & Yacobucci, K. “Data sharing and management snafu in 3 short acts” https://www.youtube.com/watch?v=N2zK3sAtr-4

Hanson, Karen, & Bakker, Theodora. 2014 “Informationist Services for Deafness Research: A Case Study” presented at NLM Board of Regents meeting, Feb 2014. http://www.slideshare.net/tabakker/informationist-services-for-deafness-research-a-case-study

McCrillis, A, Surkis, A, Vieira, D, Beam, P.S., & O'Grady, T. Survival and Success Beyond Grad School: Improving Library Services to Postdoctoral Researchers. MLA 2012

Images

cimmyt. “Planting seeds of knowledge” http://www.flickr.com/photos/cimmyt/8208414846/
wilf2. “Gummy smile” www.flickr.com/photos/wibbles/244268268
afiler. “Piggly Wiggly Flour Bag” www.flickr.com/photos/afiler/121359709
nedrichards. “Carrot Cake” http://www.flickr.com/photos/nedrichards/307600027
Svensson, Olle. “apples 2” http://www.flickr.com/photos/8070429@N06/3113672785
Fällén, Kajsa Bergman. “Oranges”. http://www.flickr.com/photos/92499343@N00/2288241903
Comendant, Quinn. “Ladies who are loves of mountain climbing” http://www.flickr.com/photos/qcom/7736318018
Mykola Swarnyk. “Simple raft” http://www.fotopedia.com/items/4tg1q9r7sq5v1-3XLChdx51D8
Bruce Guetner. “Puzzled” https://www.flickr.com/photos/10154402@N03/5322322652
Frank Kovalchek. “Barely balanced at the Arizona Renaissance Fair” https://www.flickr.com/photos/72213316@N00/5531453728
dvs. “Clark Brook Trail Hike” https://www.flickr.com/photos/dvs/3904827456
Kahunapule Michael Johnson. “Green mountains and forest” https://www.flickr.com/photos/kahunapulei/12308972825
shankar s. “I have to turn left onto the bridge now. https://www.flickr.com/photos/shankaronline/11967184863
FEN. “Young man” http://openclipart.org/detail/1169/young-man-by-fen
Don Lavange “Glass of Ayinger” https://www.flickr.com/photos/wickenden/1104589745
Thank you!

Karen Hanson
Knowledge Systems Librarian
karen.hanson@med.nyu.edu