

University of Massachusetts Medical School

eScholarship@UMMS

---

University of Massachusetts and New England  
Area Librarian e-Science Symposium

2014 e-Science Symposium

---

Apr 9th, 12:00 AM

## 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](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](#).

---

### Repository Citation

Hanson, K. L. (2014). Tales from a Data Management Survivalist: Skills Honed in the Wilderness. *University of Massachusetts and New England Area Librarian e-Science Symposium*. <https://doi.org/10.13028/yqge-kx41>. Retrieved from [https://escholarship.umassmed.edu/escience\\_symposium/2014/program/4](https://escholarship.umassmed.edu/escience_symposium/2014/program/4)

Creative Commons License

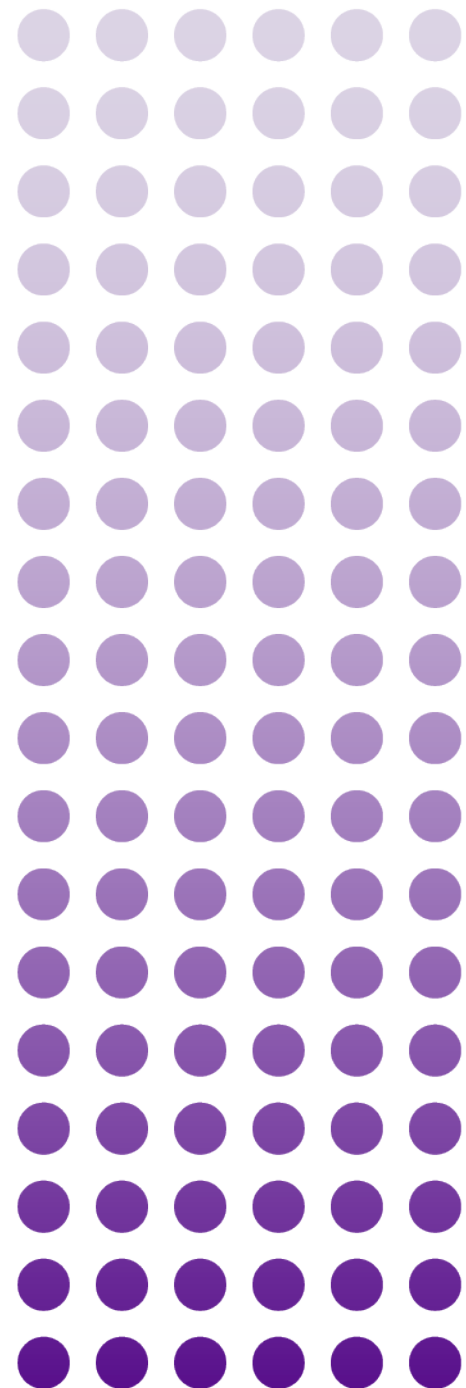


This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](#). This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto: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



me

*(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>



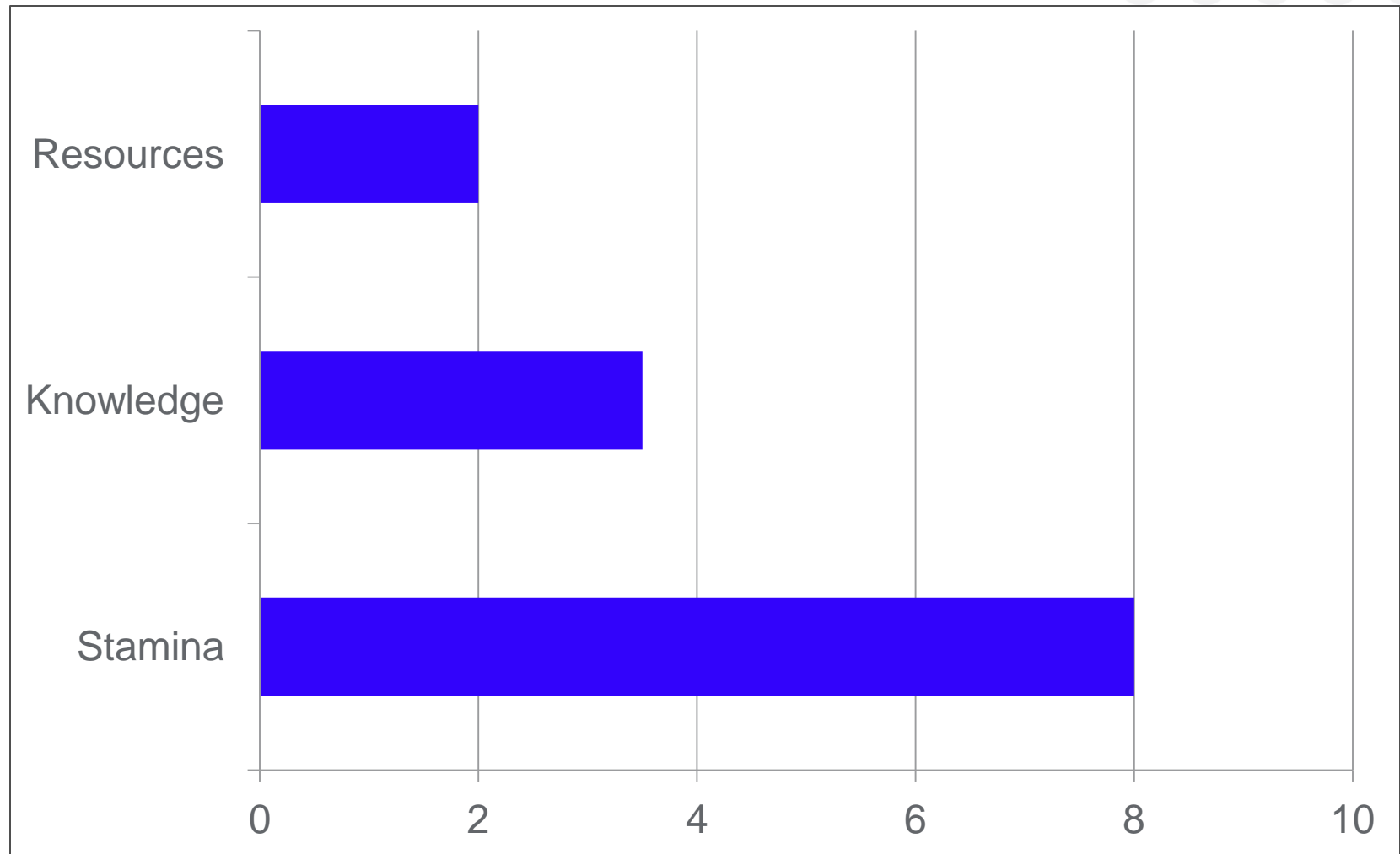
# Data services: where to start?



# Naked and afraid in the data wilderness



# Library's data strengths (2011)



- Naked and Afraid
- **Dropped in the jungle**
- Honing our survival skills
- Paddling down the river
- Lessons learned





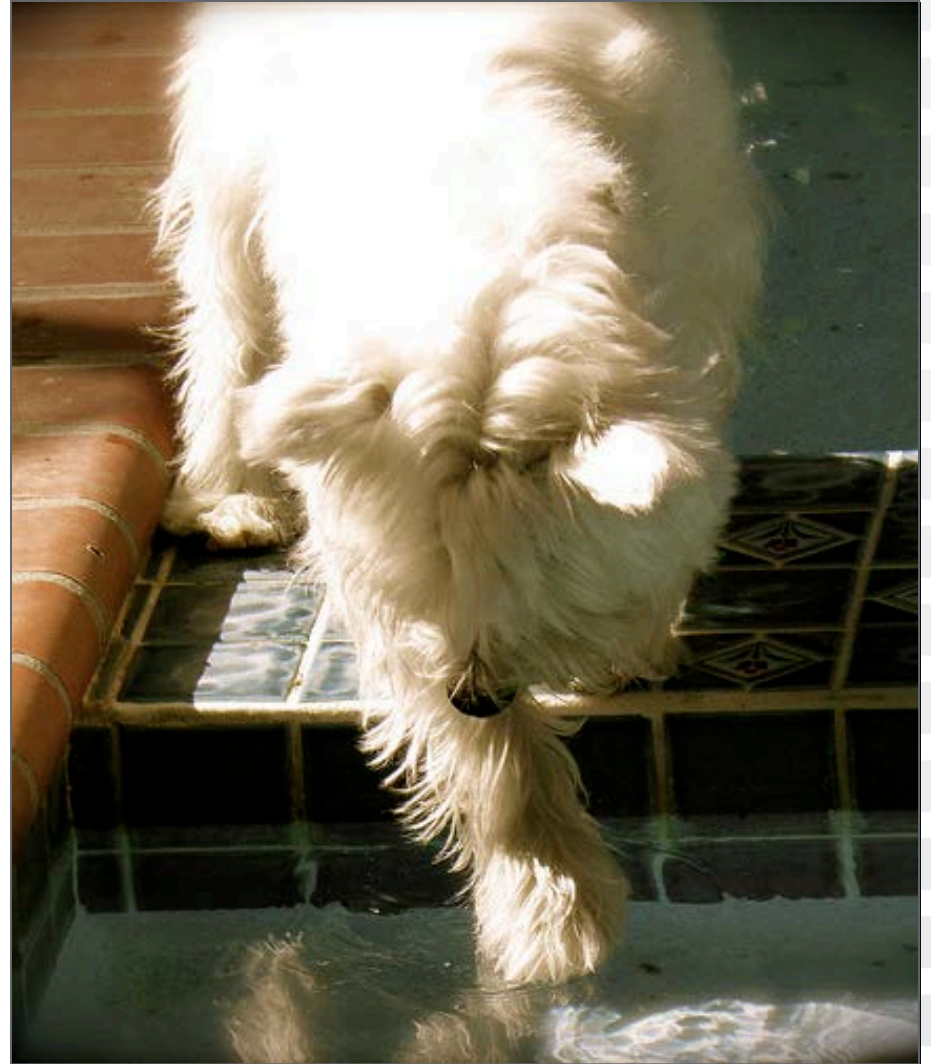
# Environmental scan

- 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!





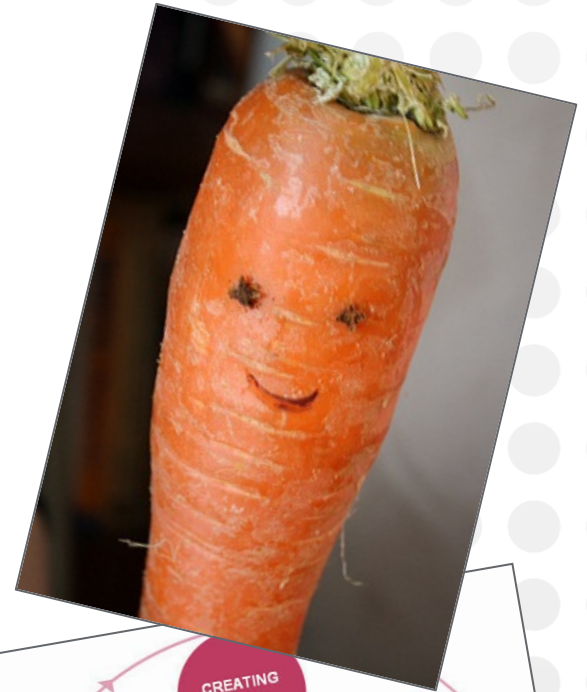
# Creating an opportunity

- 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



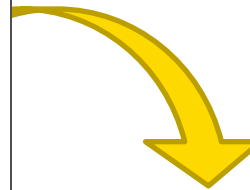
# Class features: Scare tactics

## Government mandates timeline

2003: NIH adopted a data sharing policy.

2008: NIH implements the Public Access Policy

*(still no teeth, but young yet)*



**NYU** HEALTH  
SCIENCES  
LIBRARY

## Government mandates timeline

2011: NSF made data management plans a requirement

2013: NIH Public Access Policy... now with teeth



**NYU** HEALTH  
SCIENCES  
LIBRARY

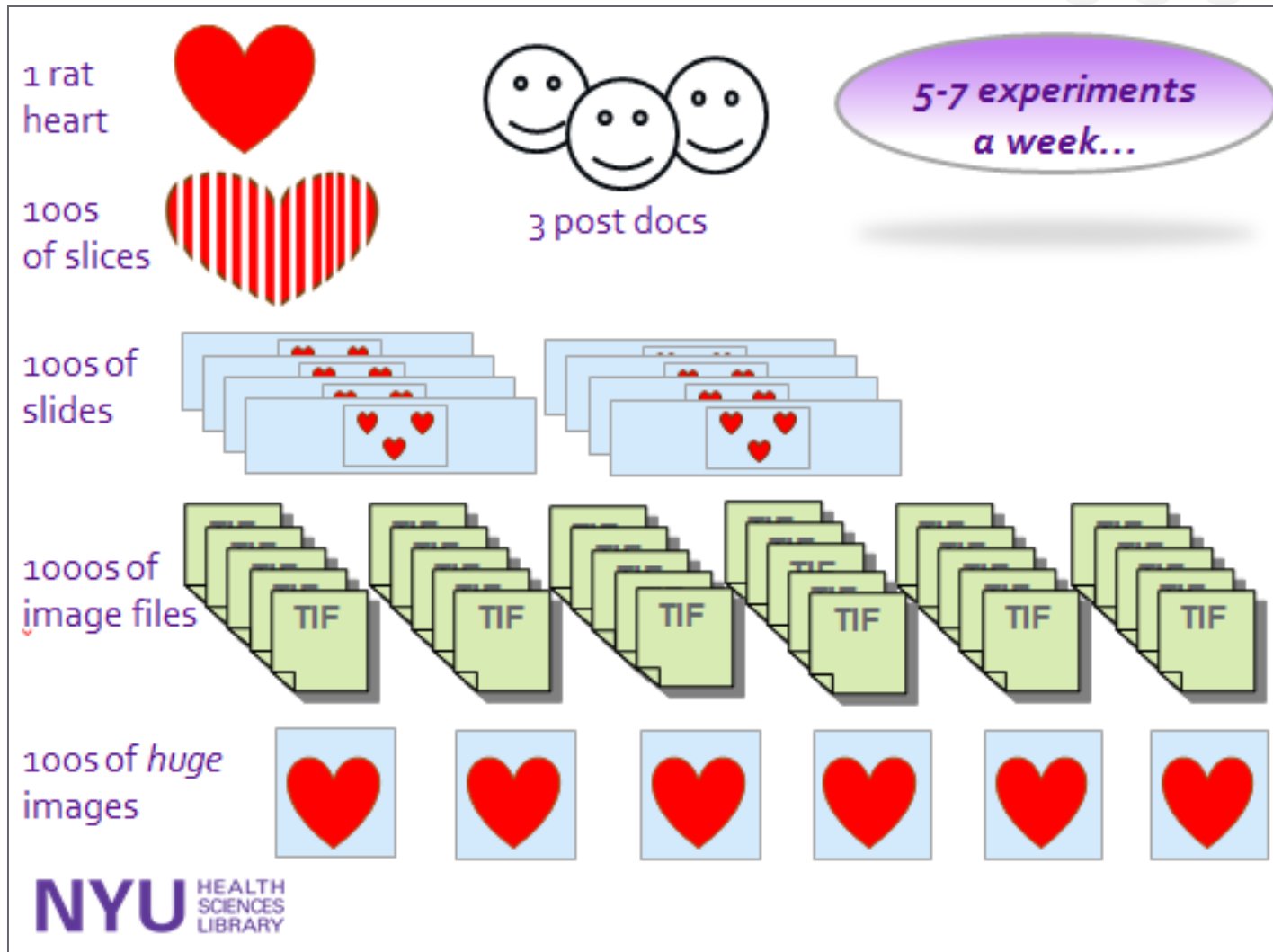


# Class features: Horror stories

*“There were 60 children in the study. The files were by accident duplicated between the upper and lower halves of the database. The files for the first 30 children in the data set were identical and in the same order with the ages for the first and 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/>

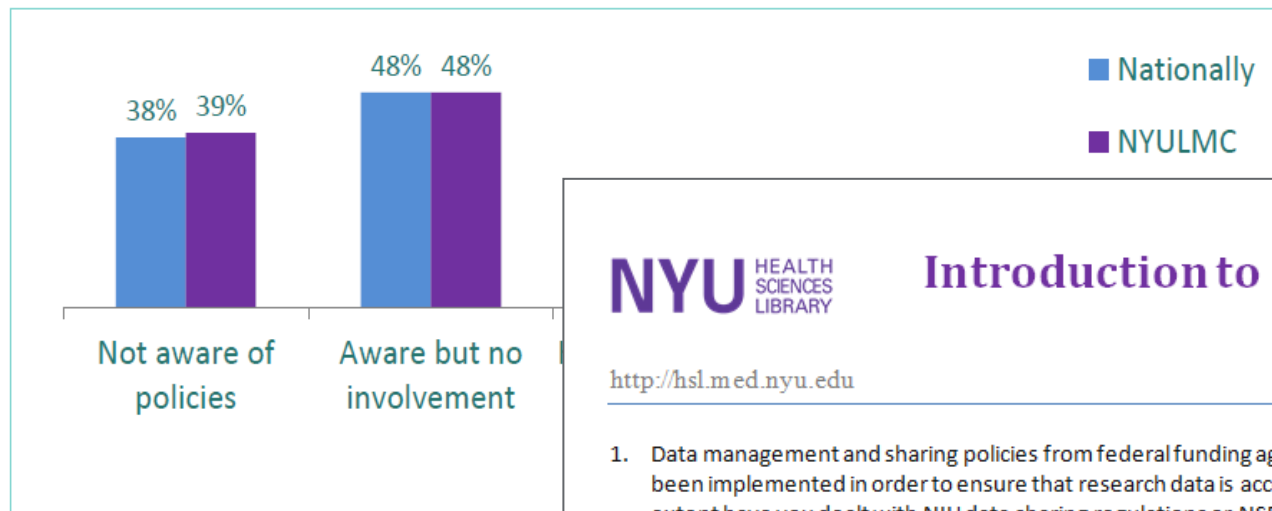
# Class features: Real world examples



# 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?



**NYU** HEALTH  
SCIENCES  
LIBRARY

## Introduction to Data Management *Pre-class Survey*

<http://hsl.med.nyu.edu>

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

# Class features: Chilling tales from our own lives



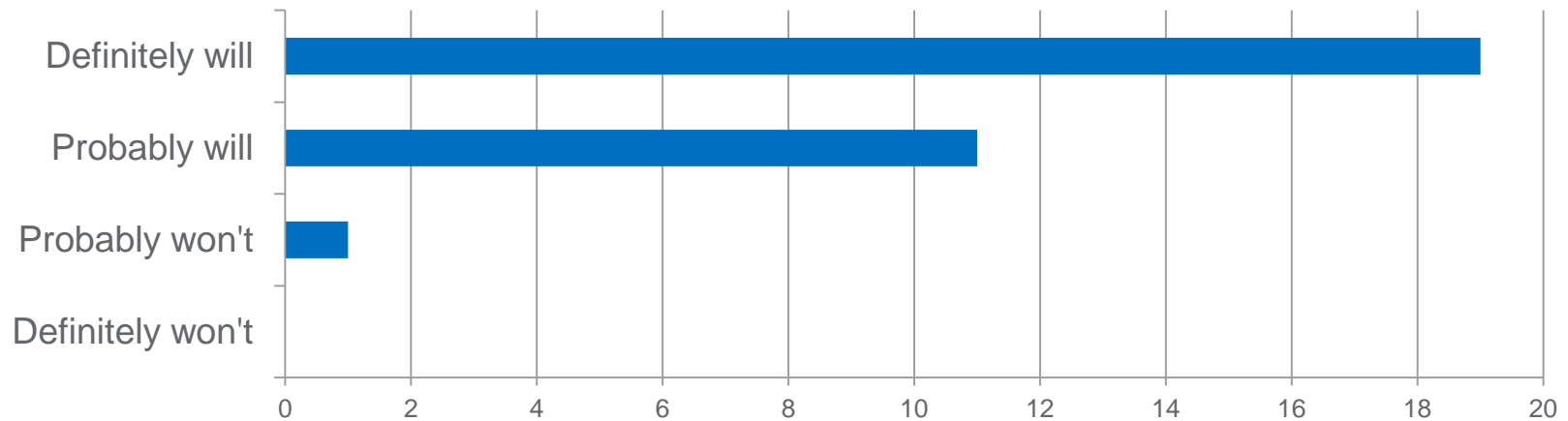
# Class features: Humor



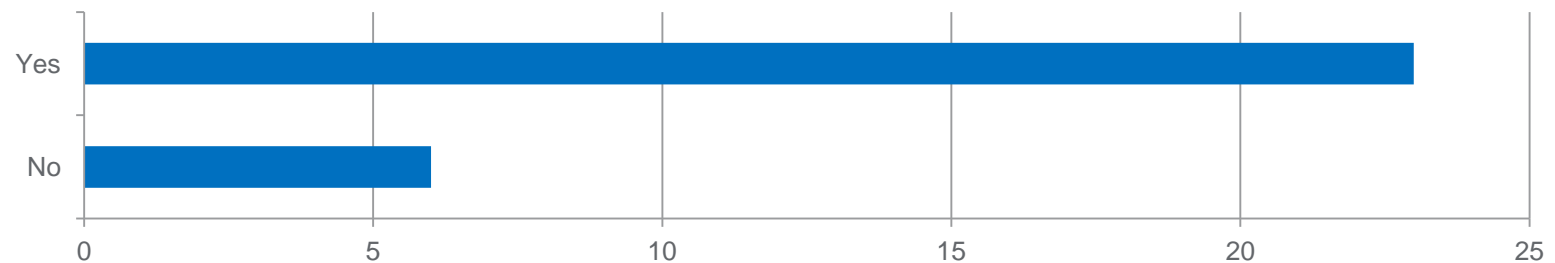


# Class evaluation

*Will you use the topics covered in your work?*



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



# 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?



me



# Division of Knowledge Informatics (DKI)





# Funding announcement

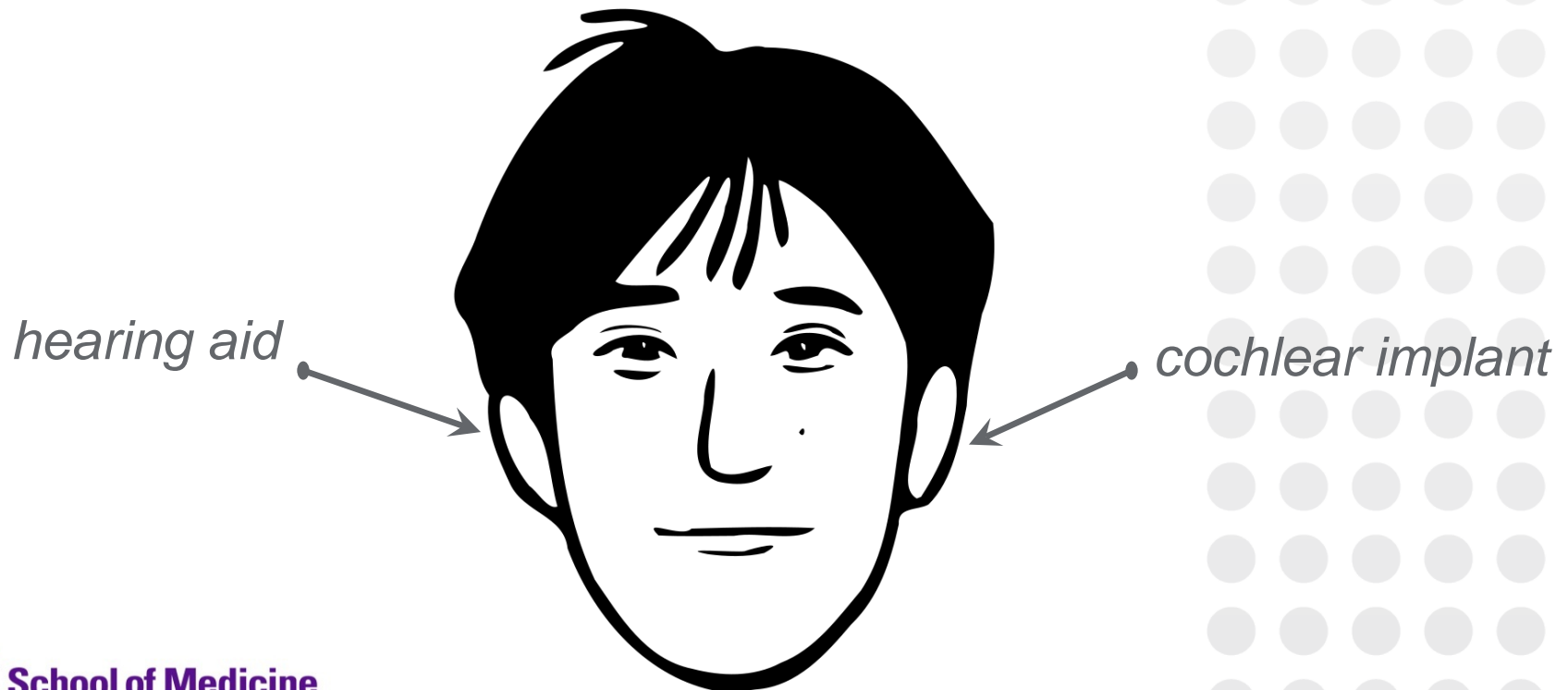
*NLM Administrative Supplements for  
Informationist Services in NIH-funded Research  
Projects*



# The grant

*“Clinical Management of Cochlear Implant Patients with Contralateral Hearing Aids”*

Mario Svirsky & Arlene Neuman

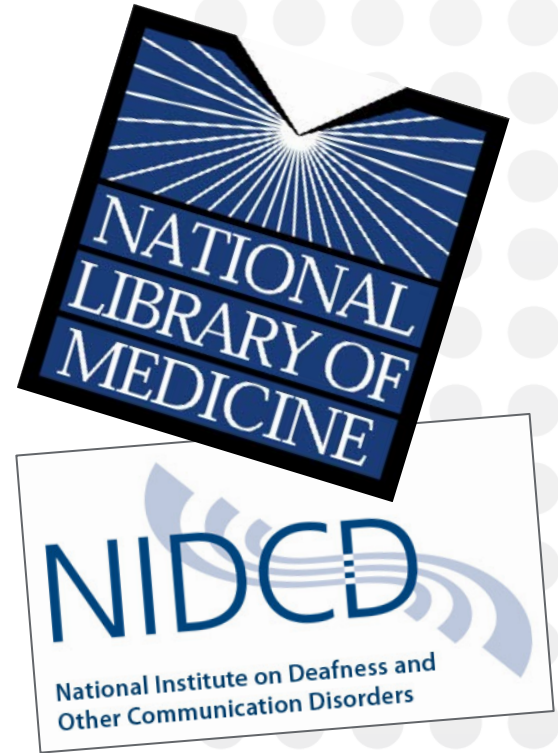


# The informationist supplement

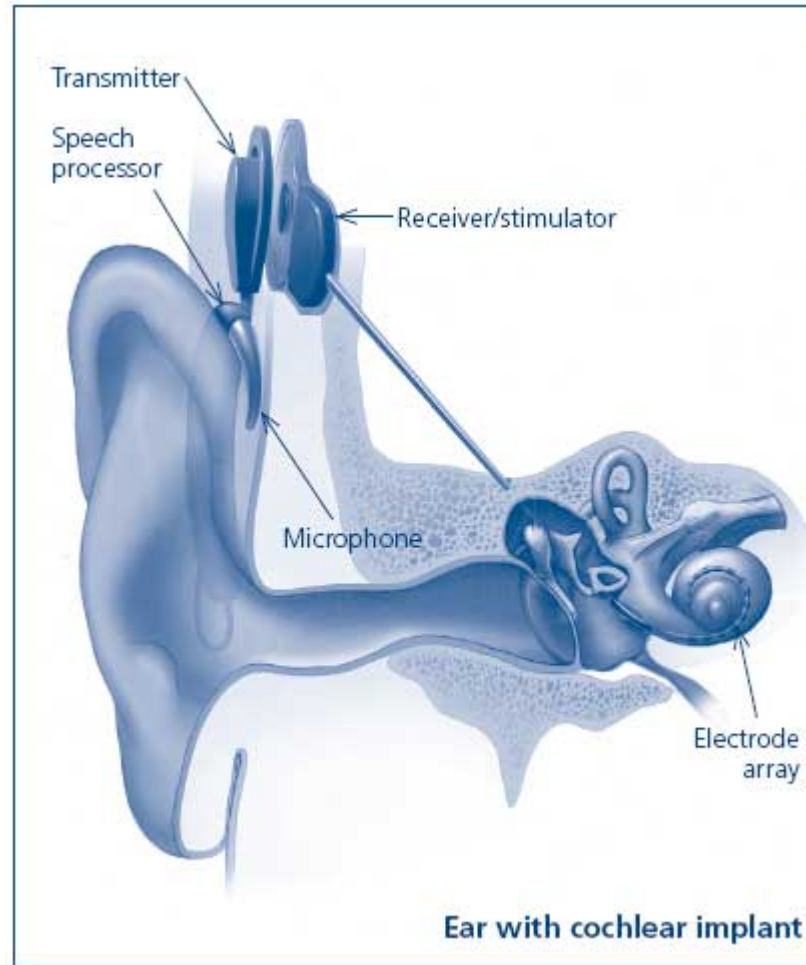
- Data model / database
- Data entry tool
- Refine reporting queries
- Query tool

## Informationists:

- Theodora: data modeling
- Me: database programming, application design

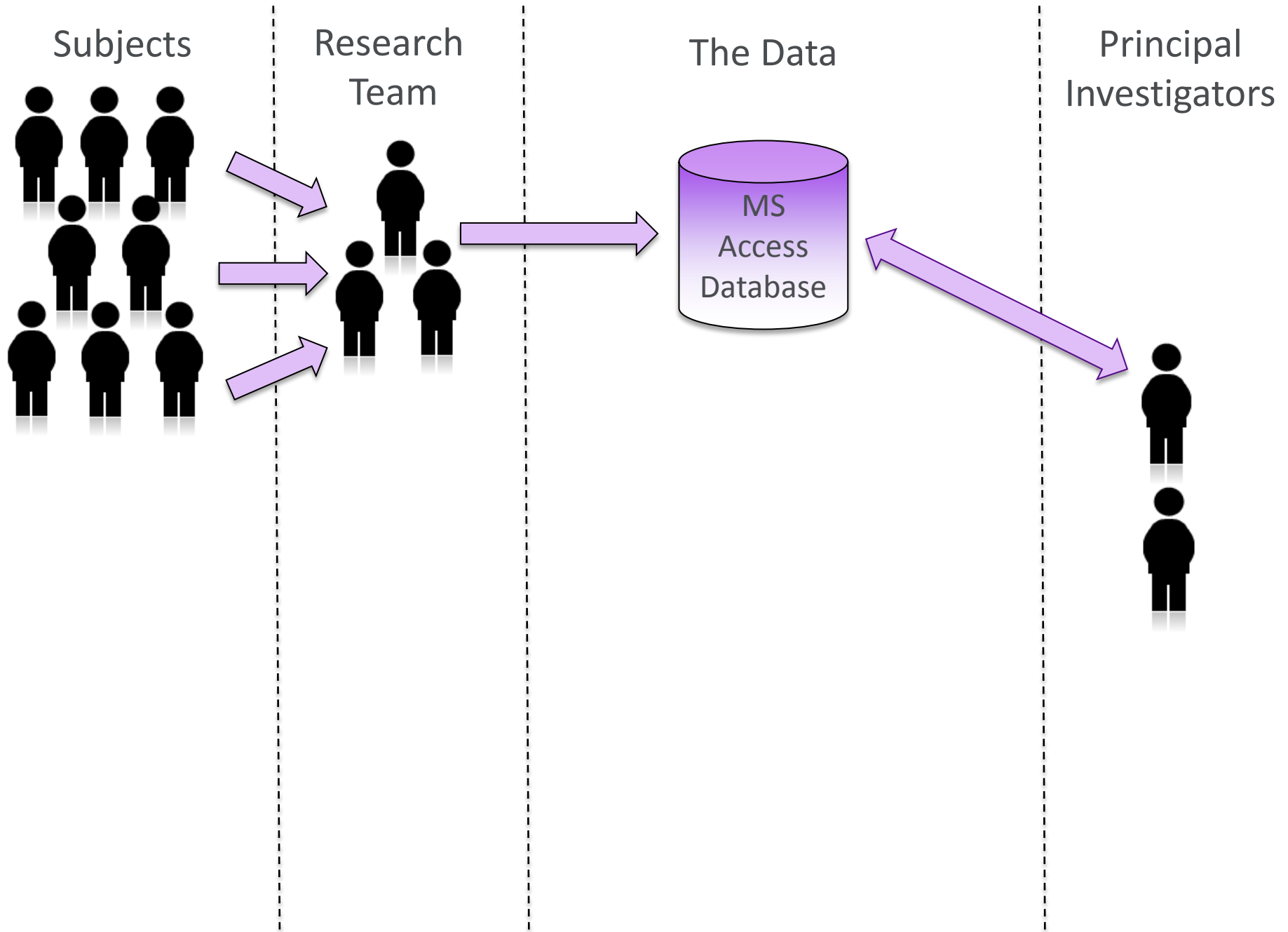


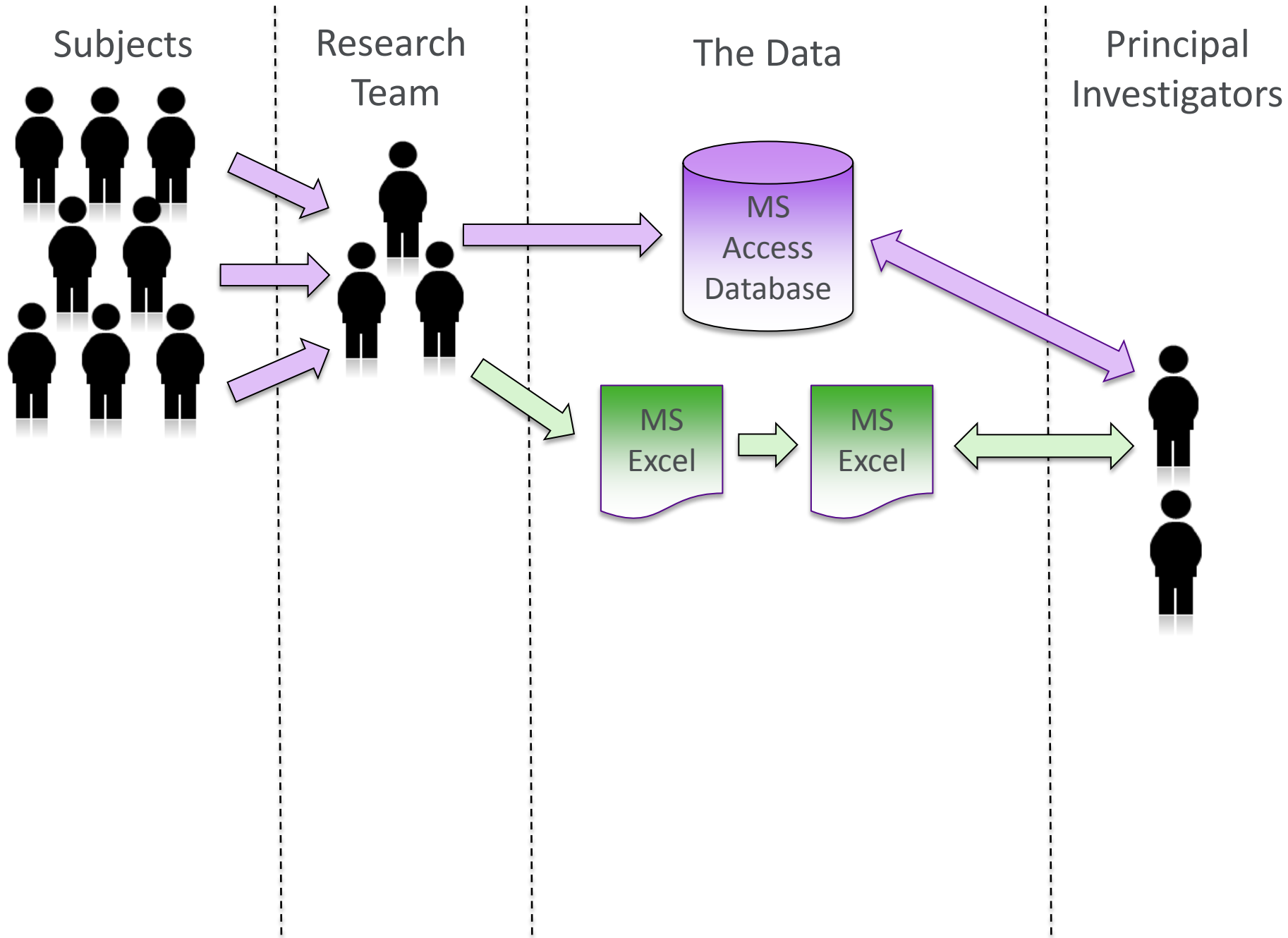
# Domain knowledge

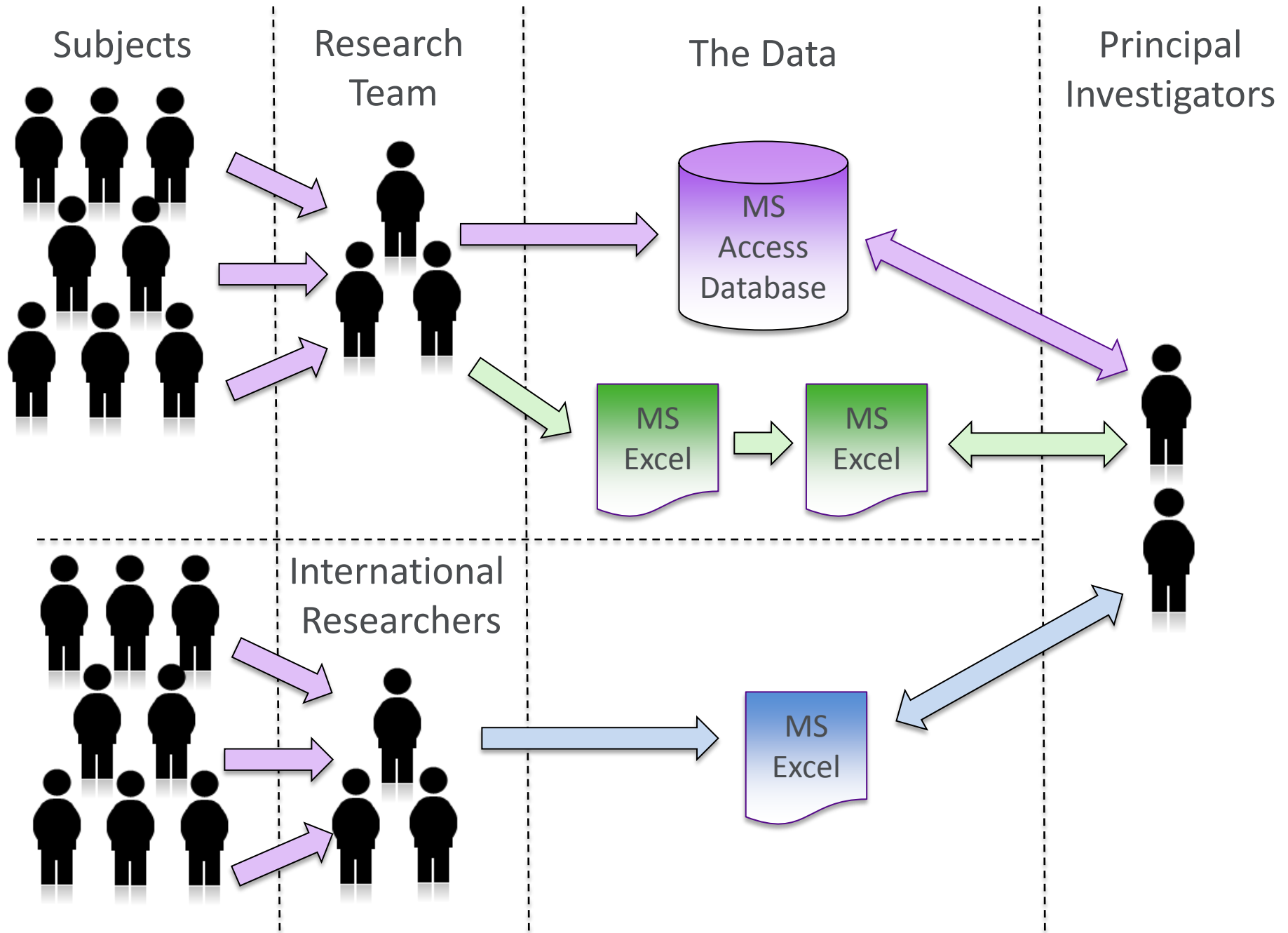


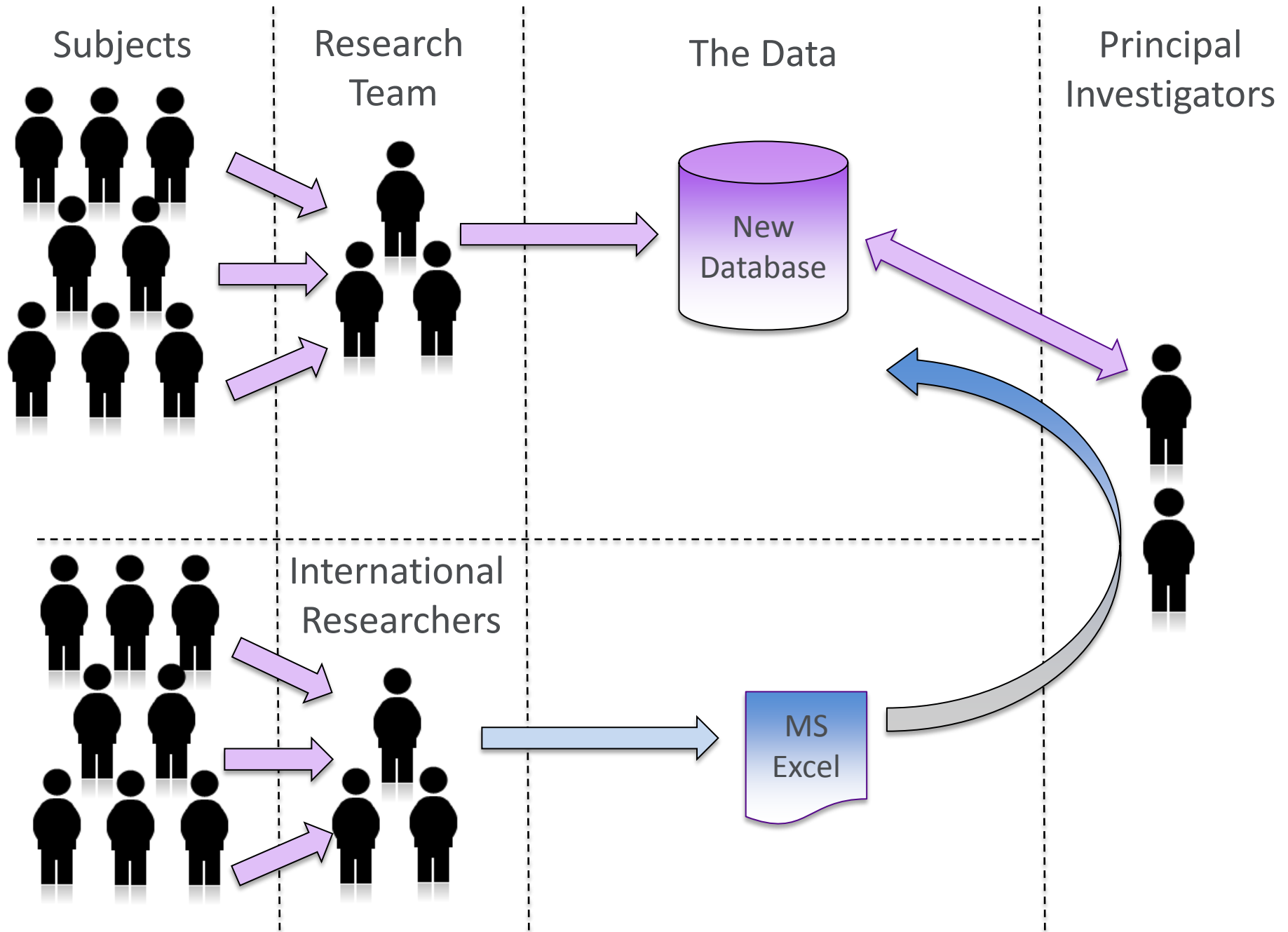
NIH Medical Arts





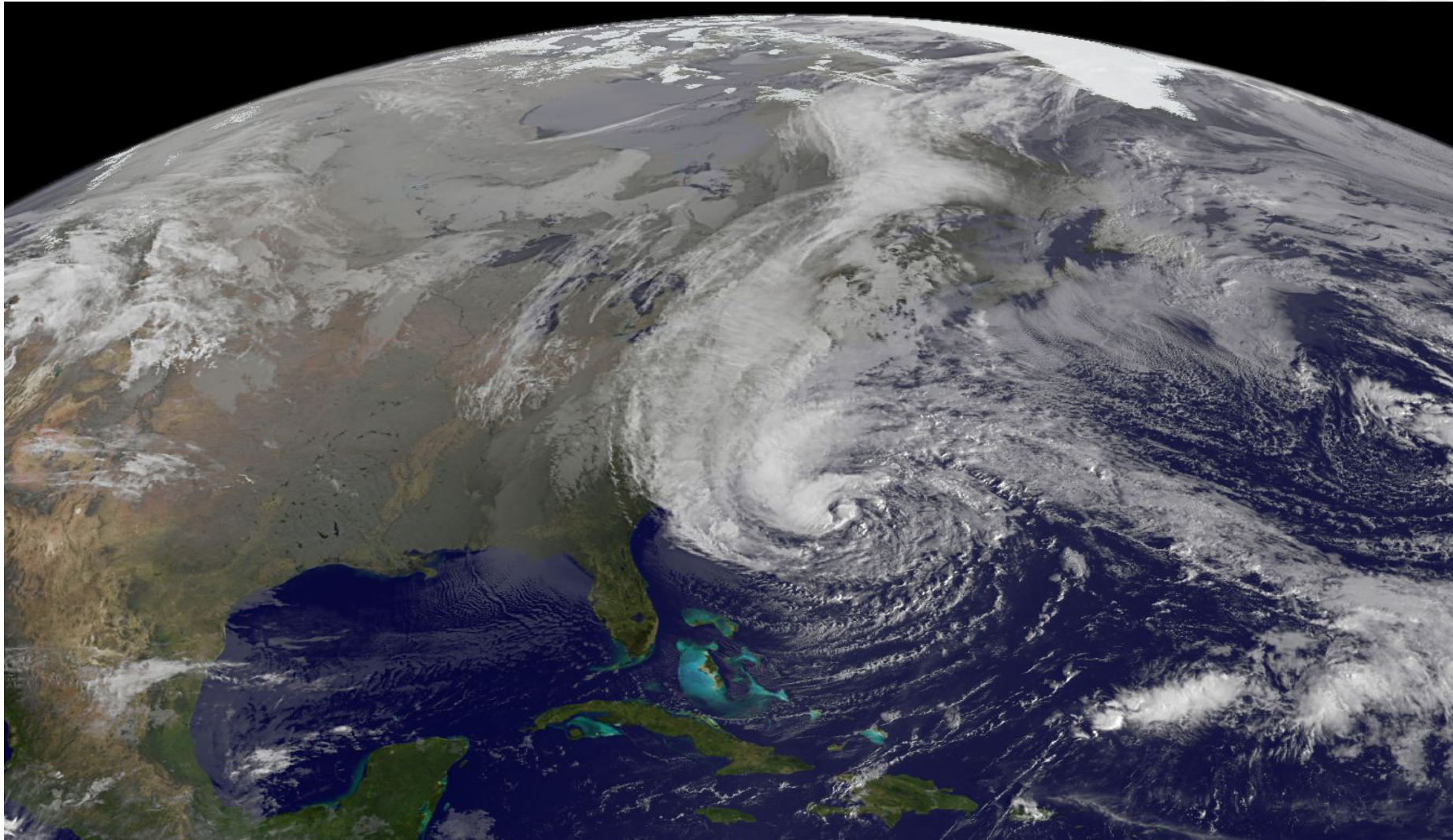








# October 2012: Hurricane Sandy



# Before



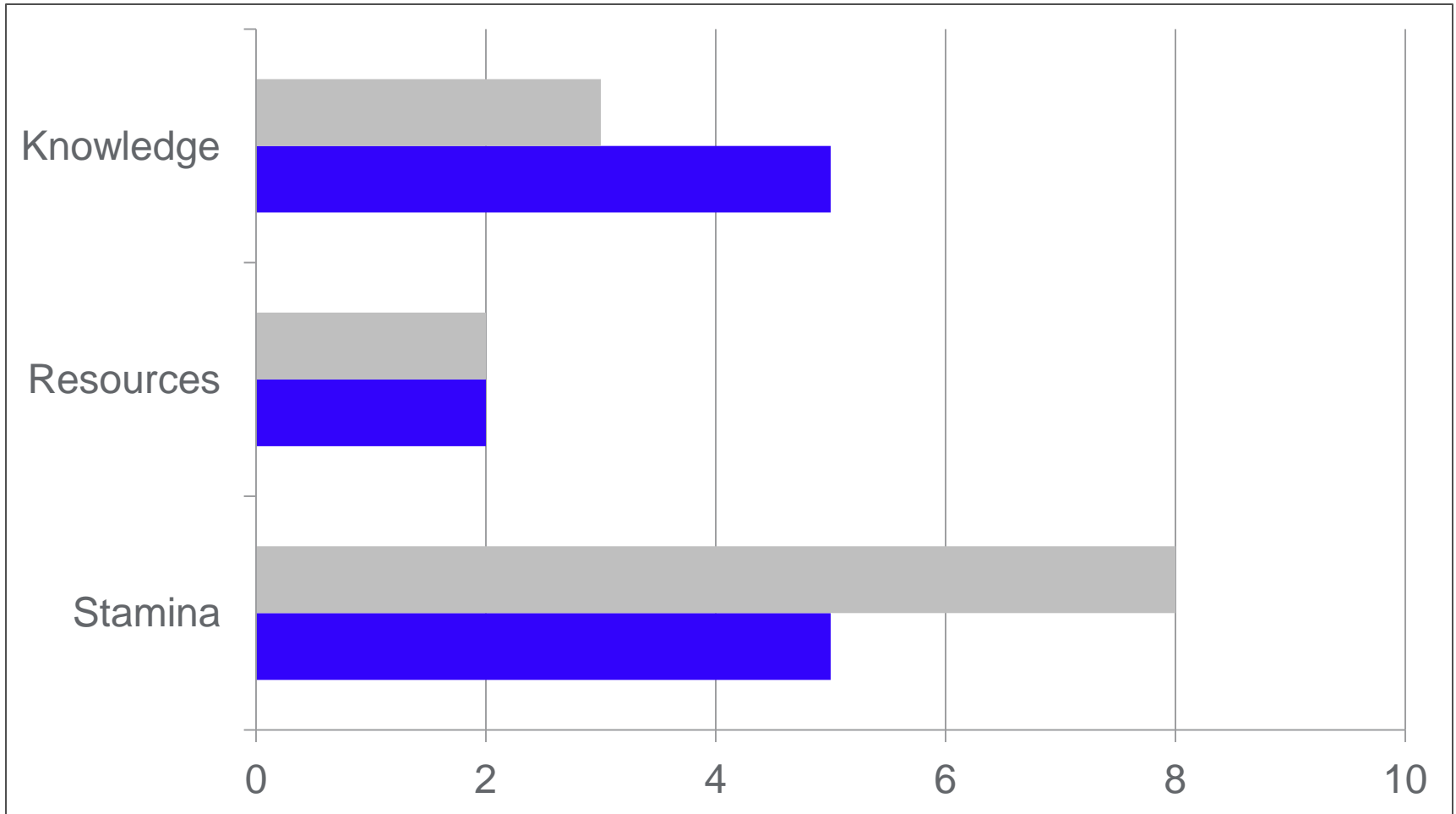
server room



# After



# Taking one on the chin



# Naked and afraid





# A glimmer of hope



The video player shows a 3D animated panda character standing in a green field under a blue sky with a large white moon. The panda has a surprised expression. In the bottom left corner of the video frame, the word "xtranormal" is visible. The video progress bar shows 2:08 / 4:40.

**2:08 / 4:40**

[Analytics](#) [Video Manager](#)

 **Data Sharing and Management Snafu in 3 Short A...**

 **NYU Health Sciences Library** · 14 videos

[Channel settings](#)

**20,010** 

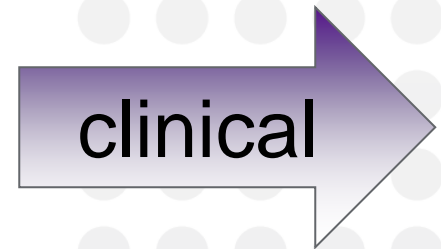
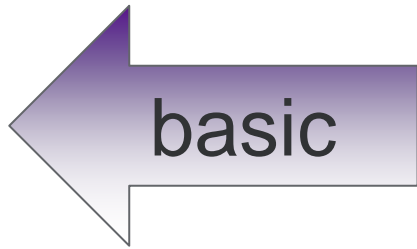
 175  2

# Early 2013



- Naked and Afraid
- Dropped in the jungle
- Honing our survival skills
- **Paddling down the river**
- Lessons learned

# A fork in the river



# 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 st

- Dublin Core
    - Designed to be generic/flexible
    - Usually stored in XML
      - e.g. <dc:creator>Harrison, Karen L.</dc:creator>
    - 15 fields
- Contributor, Coverage, Creator, D



# 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

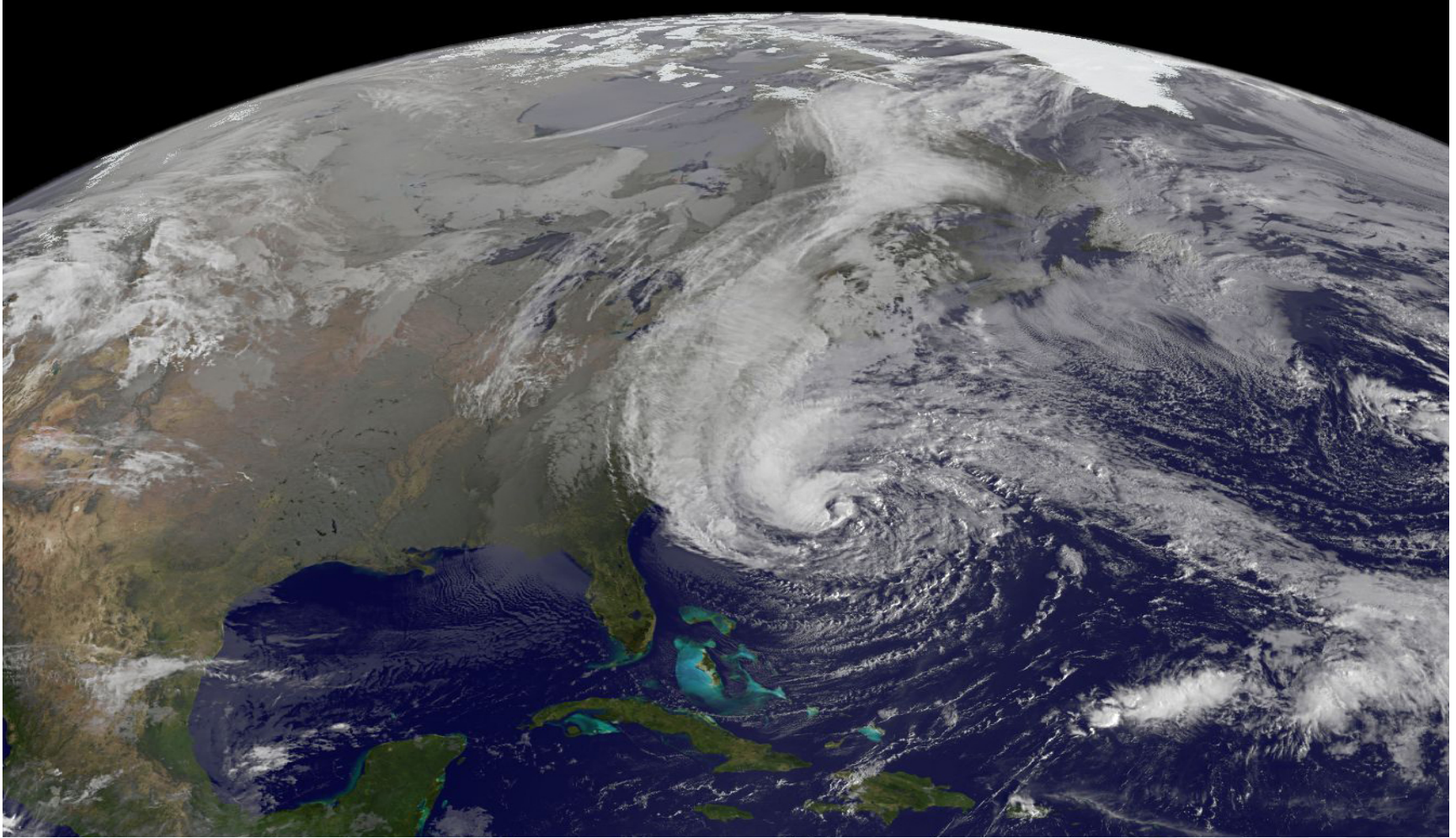


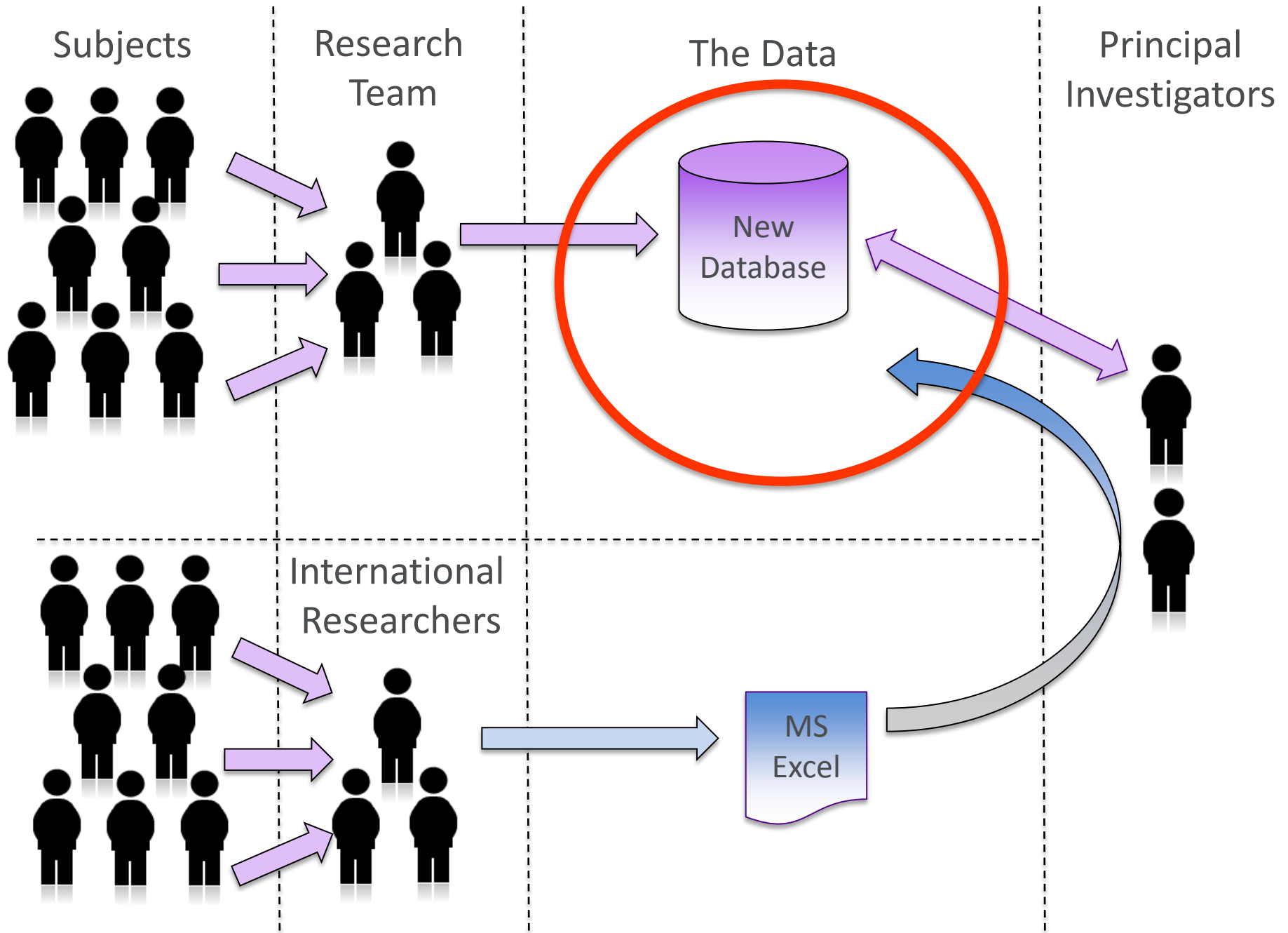
How to avoid a data  
management nightmare

NYU Health Sciences Library

Karen Hanson | Kevin Read | Alisa Surkis

# Meanwhile, the informationist project





# Tool evaluation






# Will we ever get this thing started?




# Original data entry tool

**DAT\_Subject Form**

Subject ID:  Last Name:  FirstName:  

Subject Data Parent / Guardian Left Implant Right Implant Hearing Aids Test Scores Other Info

Implant ID:  

Ear:

Implant Device:

Implant Date:

No. of Electrode:


Stimulation Date:

Active:

**Processor Detail**


Implant ID:  Processor:

Processor ID:  Processor Date:

Active 

**Strategy Detail**

Processor ID:  Strategy ID:

Strategy:  

Strategy Date:

Active

Record:  of 1

Record:  of 1

Record:  of 1

# Tool evaluation

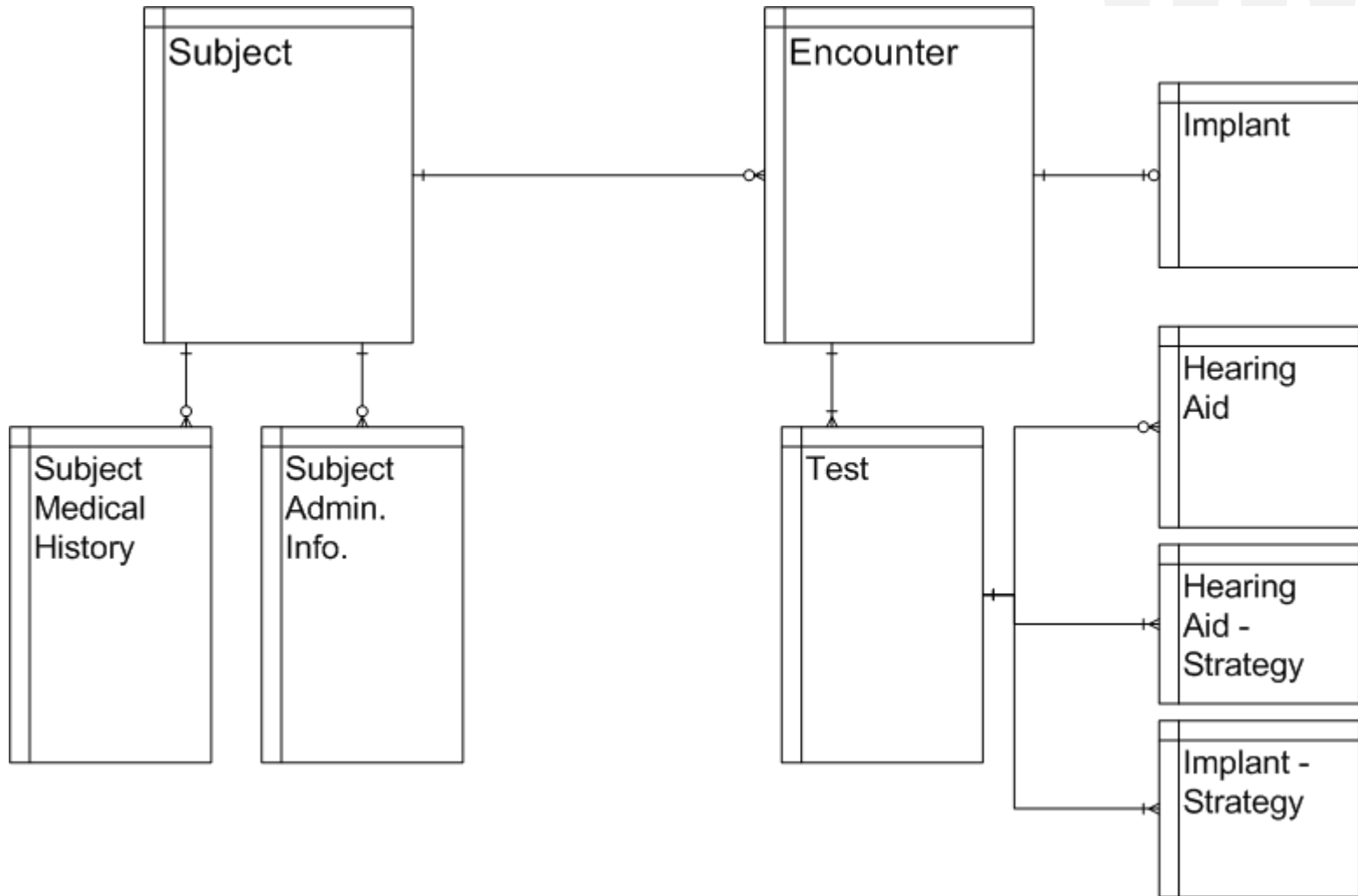


OK, we're in it for the long haul





# A unified model



# Cleaner data entry

Visit details - Auditory Research

http://audstudies.med.nyu.edu

## Auditory research tool

hansok01 is logged in | [log out](#)

Manage subjects | **Manage visit** | Create reports | Configure lists | Configure users

Editing visit:  
Marc Botton  
12/01/2013

**Visit details**  
Assign tests  
Score tests

### Visit details

Cancel Save changes

General details

Subject: 3 Marc Botton

Visit date: 12/01/2013 Calculated age: 36.123

Communication mode: Oral/Cued

Site: Select site

Cochlear implant details

Left ear has CI  Right ear has CI

Implant device: Nucleus 22 Select implant device

Implant date: 12/01/2013 / /

# electrodes: 22

Stimulation type: Select stimulation type Select stimulation type

Stimulation date: 12/01/2013 / /

Insertion angle: 22

Insert technique: Select insertion technique Select enlargement technique

# Validation, autocomplete, audit

### Score tests

Filter list:

Test Code	Description	Score
<u>23</u>	AudioMetric 500 Aided Left Ear	1000 *
<u>26</u>	AudioMetric 500 Aided Right Ear	100
<u>27</u>	AudioMetric 4000 UnAided SF/WT	98

★ The highlighted score falls outside the accepted range (0-100).  
Please verify before saving.

### Cause of hearing loss

Aetiology

Otitis media

M

Aetiology notes

Meningioma

Meningitis

Multiple sclerosis

Validation status

Visit record status

# Built-in and custom reporting

Enter report name or ID   [Create new report](#)

## Built in reports

Report ID	Report Name	Last Run	
<u>5</u>	<u>Intelligibility report</u>	<u>09/30/2013</u>	<u>run   duplicate</u>
<u>4</u>	<u>Expressive report</u>	<u>12/01/2013</u>	<u>run   duplicate</u>
<u>3</u>	<u>Bimodal report</u>	<u>12/01/2013</u>	<u>run   duplicate</u>
<u>2</u>	<u>General report</u>		
<u>1</u>	<u>CI report</u>		

## Custom reports

Report ID	Report Name
<u>8</u>	<u>International subje</u>
<u>9</u>	<u>CI subgroup</u>
<u>12</u>	<u>Bimodal recent</u>
<u>16</u>	<u>Potential subjects</u>
<u>21</u>	<u>Test report 1</u>

### Bimodal audiometric report (built in)

Score

Select column  Type  Enter term for filter

[Export view to Excel](#)

Subject	BirthDat	LeftInitImpDat	RightInitImpDat	TestDate	LeftTestDe	RightTestDe	Scor	Test
158	2012-01-0	2008-04-03		2009-03-0	No Aid	No Aid	75	AudioMetric 2000 UnAided Right
158	2012-01-0	2008-04-03		2008-02-2	No Aid	No Aid	85	AudioMetric 4000 UnAided Left
158	2012-01-0	2008-04-03		2008-02-2	No Aid	No Aid	45	AudioMetric 250 UnAided Right
158	2012-01-0	2008-04-03		2008-02-2	No Aid	No Aid	65	AudioMetric 1000 UnAided Left
158	2012-01-0	2008-04-03		2008-02-2	No Aid	No Aid	55	AudioMetric 500 UnAided Right
158	2012-01-0	2008-04-03		2009-03-0	No Aid	No Aid	100	AudioMetric 4000 UnAided Right
158	2012-01-0	2008-04-03		2009-03-0	No Aid	No Aid	65	AudioMetric 1000 UnAided Right
158	2012-01-0	2008-04-03		2008-02-2	No Aid	No Aid	50	AudioMetric 250 UnAided Left Ear
158	2012-01-0	2008-04-03		2008-02-2	No Aid	No Aid	90	AudioMetric 2000 UnAided Left
158	2012-01-0	2008-04-03		2008-02-2	No Aid	No Aid	60	AudioMetric 1000 UnAided Right
158	2012-01-0	2008-04-03		2009-03-0	No Aid	No Aid	50	AudioMetric 250 UnAided Right
158	2012-01-0	2008-04-03		2009-03-0	No Aid	No Aid	55	AudioMetric 500 UnAided Right
158	2012-01-0	2008-04-03		2008-02-2	No Aid	No Aid	90	AudioMetric 4000 UnAided Right



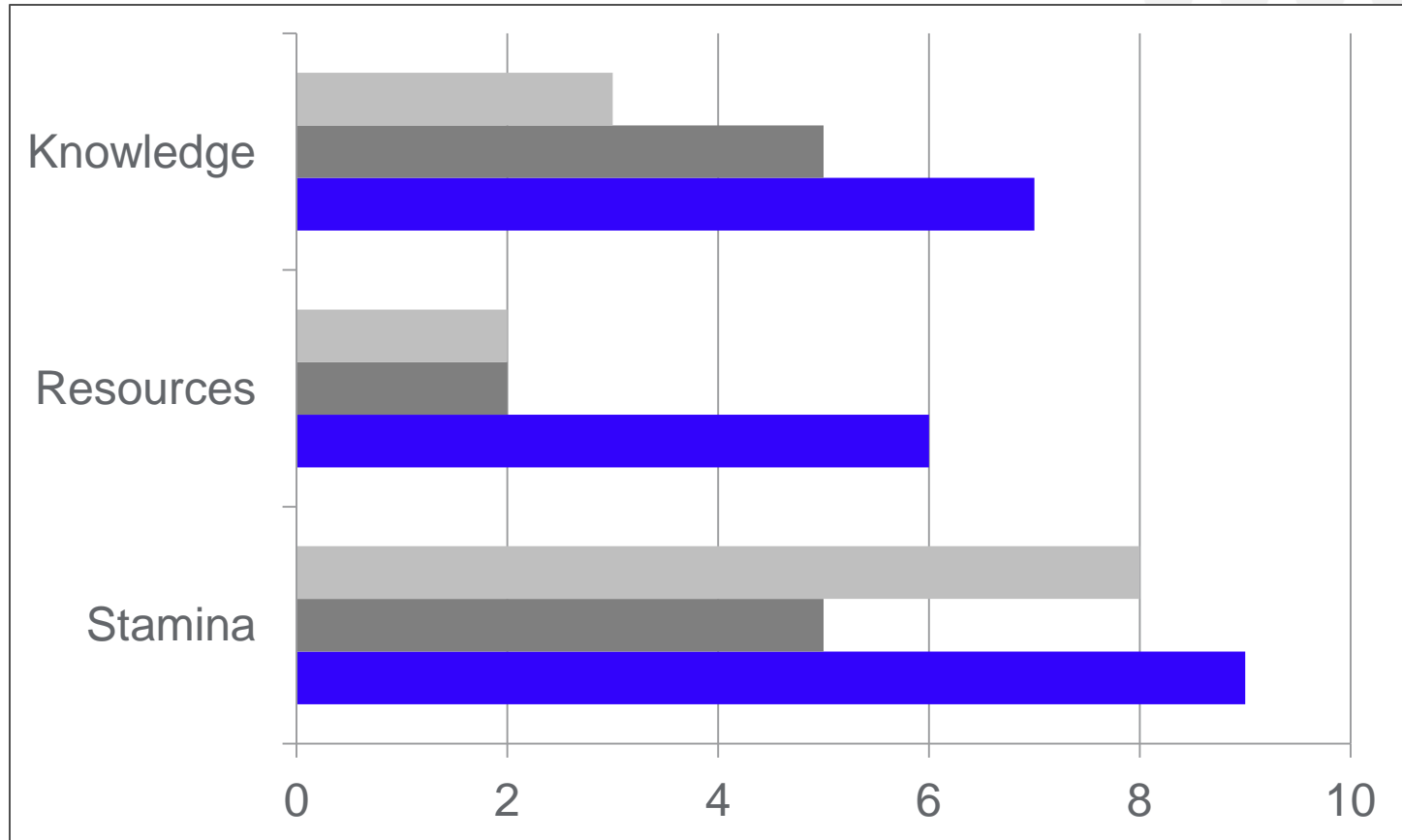
# Informationist supplement – take aways

- 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



# 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

ACRL. Changing Roles of Academic and Research Libraries. 2006 <http://www.ala.org/acrl/issues/value/changingroles>

Gaudette, G. Presentation at UMass' 2012 New England eScience Symposium. (Cardiology example)  
[http://escholarship.umassmed.edu/escience\\_symposium/2012/program/9/](http://escholarship.umassmed.edu/escience_symposium/2012/program/9/)

Hanson, K, Surkis, A, & Read, K. "Introduction to Data Management" [http://hslguides.med.nyu.edu/data\\_management](http://hslguides.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

*Retraction: Vitamin C and asthma in children: modification of the effect by age, exposure to dampness and the severity of asthma*. 2012. PubMed Central. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3320558>

# Images

UK Data Archive. Data lifecycle image. <http://www.data-archive.ac.uk/create-manage/life-cycle>

purpleapple428. "Let's Go Swimming" <http://www.flickr.com/photos/purpleapple428/5452339625/>

cimmyt. "Planting seeds of knowledge" <http://www.flickr.com/photos/cimmyt/8208414846/>

wilf2. "Gummy smile" [www.flickr.com/photos/wibbles/244268268](http://www.flickr.com/photos/wibbles/244268268)

outcast104. "Vampire weekend" [www.flickr.com/photos/outcast104/2011632229](http://www.flickr.com/photos/outcast104/2011632229)

afiler. "Piggly Wiggly Flour Bag" [www.flickr.com/photos/afiler/121359709](http://www.flickr.com/photos/afiler/121359709)

Mel B. "Oil pour" 2008. <http://www.flickr.com/photos/42dreams/2452877486>

psrobin. "Baking Powder Still Life" [www.flickr.com/photos/psrobin/5092598788](http://www.flickr.com/photos/psrobin/5092598788)

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>

Liv Unni Sødem. "Caiman attack in Brazil" <https://www.flickr.com/photos/livunni/3310847659>

Matthew Hutchinson. "Fork" <https://www.flickr.com/photos/hiddenloop/7945924094>

Mykola Swarnyk. "Simple raft" <http://www.fotopedia.com/items/4tg1q9r7sq5v1-3XLChdx51D8>

Luke Jones. "Jungle" <https://www.flickr.com/photos/befuddledsenses/1334533356>

Jelene Morris. "my tank just cleaned" <https://www.flickr.com/photos/jelene/2634767417>

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>

ooh.ooh. "handshake 1" <http://www.fotopedia.com/items/flickr-1350774613>

dvs. "Clark Brook Trail Hike" <https://www.flickr.com/photos/dvs/3904827456>

Kahunapule Michael Johnson. "Green mountains and forest" <https://www.flickr.com/photos/kahunapulej/12308972825>

Jesse the Traveler. "Steamy Jungle Trail" <https://www.flickr.com/photos/jesseslife/310218074>

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>

Christian F. Burprich. Man, silhouette, user icon. [https://www.iconfinder.com/icons/16992/man\\_silhouette\\_user\\_icon#size=128](https://www.iconfinder.com/icons/16992/man_silhouette_user_icon#size=128)

FileSquare. "Excel icon" [https://www.iconfinder.com/icons/79354/excel\\_icon#size=128](https://www.iconfinder.com/icons/79354/excel_icon#size=128)

Don Lavange "Glass of Ayinger" <https://www.flickr.com/photos/wickenden/1104589745>

Simmon R. Geostationary Operational Environmental Satellite 13: Hurricane Sandy. National Aeronautics and Space Administration. Oct 18, 2012. <http://earthobservatory.nasa.gov/NaturalHazards/view.php?id=79553>

Chris Walts. "Banana tree" <https://www.flickr.com/photos/crashadventures/5973206296>

NIDCD. "Cochlear implants" <http://www.nidcd.nih.gov/health/hearing/pages/coch.aspx>

**Thank you!**

Karen Hanson  
Knowledge Systems Librarian  
[karen.hanson@med.nyu.edu](mailto:karen.hanson@med.nyu.edu)

