NIH fMRI Summer Course

fMRI Data Sharing

Daniel Handwerker
Section on Functional Imaging Methods
Laboratory of Brain and Cognition
National Institute of Mental Health
U.S. Department of Health & Human Services

August 5, 2015

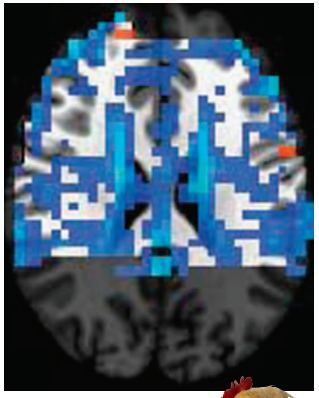
What is Data Sharing?

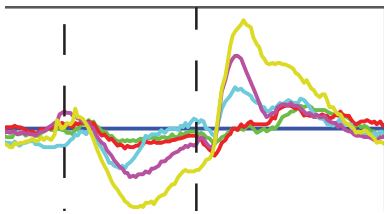
The practice of making data used in scholarly research available to other researchers

https://en.wikipedia.org/wiki/Data_sharing

What is "data"?

Time-to-peak	8.6 ± 0.9	7.6 ± 1.0	7.7 ± 1.2	7.4 ± 1.2	6.9 ± 1.0
Time-to-valley	8.3 ± 1.2	11.7 ± 1.2	11.9 ± 1.6	10.8 ± 1.8	10.5 ± 1.6

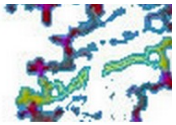




Fully prepocessed MRI data?
"Raw" MRI
Stimulus presentation code?
Behavioral data?
Data processing code?

What is "available"?

dbSNP
Short Genetic Variations



National Database for Autism Research (NDAR)

OpenfMRI

Overview

Why share data

Why most people don’t share data

Why fMRI data sharing is improving

Why share data?

Selfish reason #1

The most important person you share data with is

Yourself

The next most important people are
your current & future collaborators

A case study

(About a specific situation, but this story is VERY common)

- Re-examining data collected years ago
- fMRI data are in two different formats and spread out across several semi-arbitrary directories
- Behavioral data is on a separate computer
- Records about what happened in each scan are elsewhere

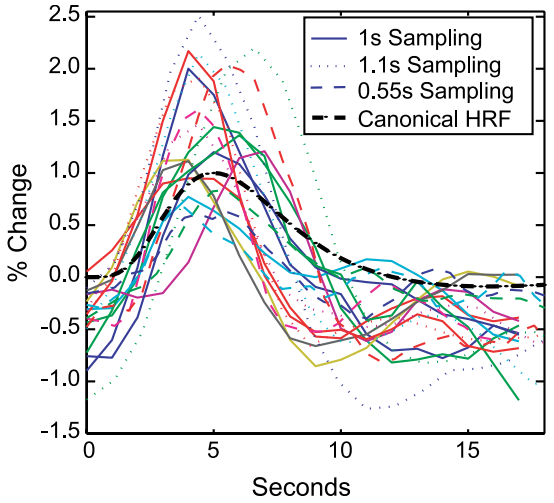
Just wrangling the data to a point where it was again useful was a big project!

If one or two key people left research lab, the data may have been un-usable

An even closer-to-home case study

Handwerker et al 2004 Neuroimage

The data to make the plots in the figures are fairly accessible to me and a few others



The full fMRI volumes are on a few external hard drives in a format that isn't used anymore

```
101/anatomy/axials.cub.gz
101/tes_data/eyerf101_correct.tes.gz
102/anatomy/axials.cub.gz
103/anatomy/axials.cub.gz
103/tes_data/eyerf_103correct.tes.gz
```

```
104/anatomy/axials.cub.gz
104/tes_data/eyerf104_correct.tes.gz
105/anatomy/axials.cub.gz
105/tes_data/eyerf105_correct.tes.gz
```

How many 5-year-old studies could be replicated from collected data to publication?

How many just published studies?

Why share data?

Selfish reason #2

Sharing Data is increasingly becoming a requirement for publication & grants

Proceedings of the National Academy of Sciences (PNAS) Editorial Policies

To allow others to replicate and build on work published in PNAS, authors must make materials, data, and associated protocols available to readers.

http://www.pnas.org/site/authors/journal.xhtml

NIH Genomic Data Sharing Policy

As a condition of grant funding, human and non-human genomic data must be submitted to an NIH-designated repository in a timely manner (with some exceptions)

http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html

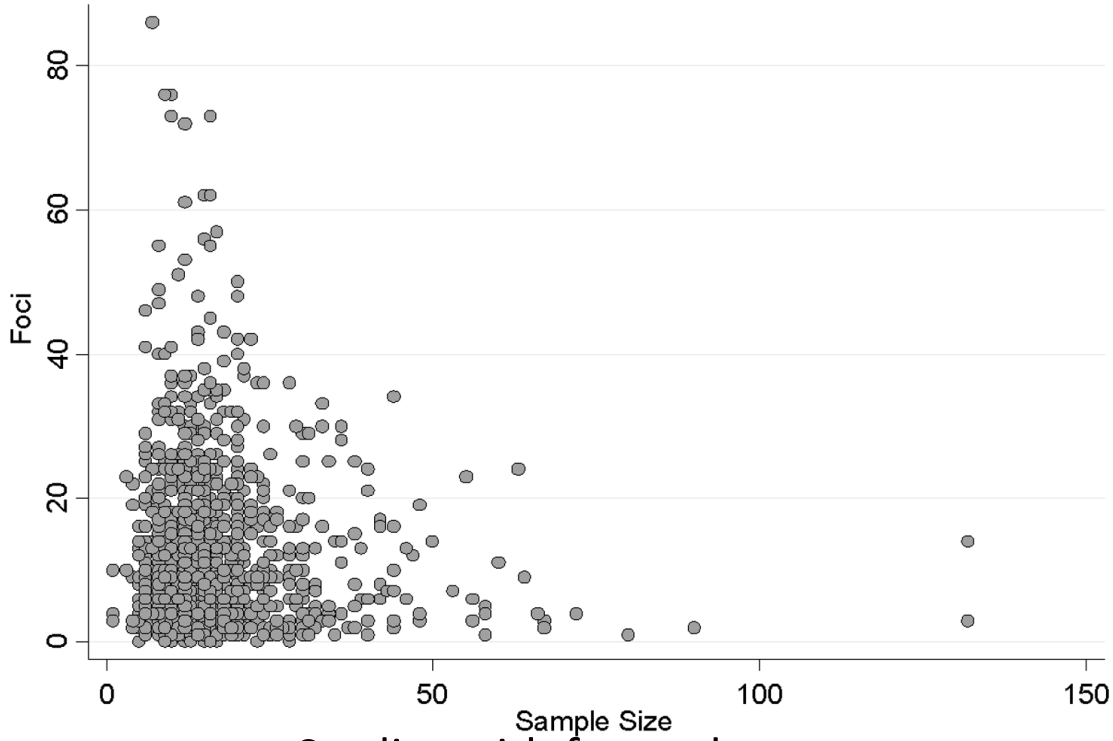
Sharing data -> Better science

A typical fMRI study

- Design a study
- Collect data from 15-30 people
- A few people in a group look at the images, time series, and statistical maps
- The same people (hopefully) write up their finding and publish
- The data are kept on a hard drive somewhere or archived

How does this typical fMRI study limit our understanding of the brain?

- Minimal replication (fMRI is expensive) or replications with slightly different results (Not necessarily bad)
- Why do results differ?
  - Data quality?
  - Experimental design decisions?
  - Sample size?
  - Analysis decisions?



Studies with fewer data sets have more foci of activation

David, Ware, et. al. 2013. Potential Reporting Bias in fMRI Studies of the Brain. PLoS ONE doi:10.1371/journal.pone.0070104

How this changes with Sharing

- Much larger sample sizes
- Better understanding of replication attempts

  If there is inconsistency, then directly compare data sets
- People without resources to collect high quality fMRI data can make important contributions to the field
  - Experts in other areas can offer new insights
  - Better ways to develop and compare analysis methods

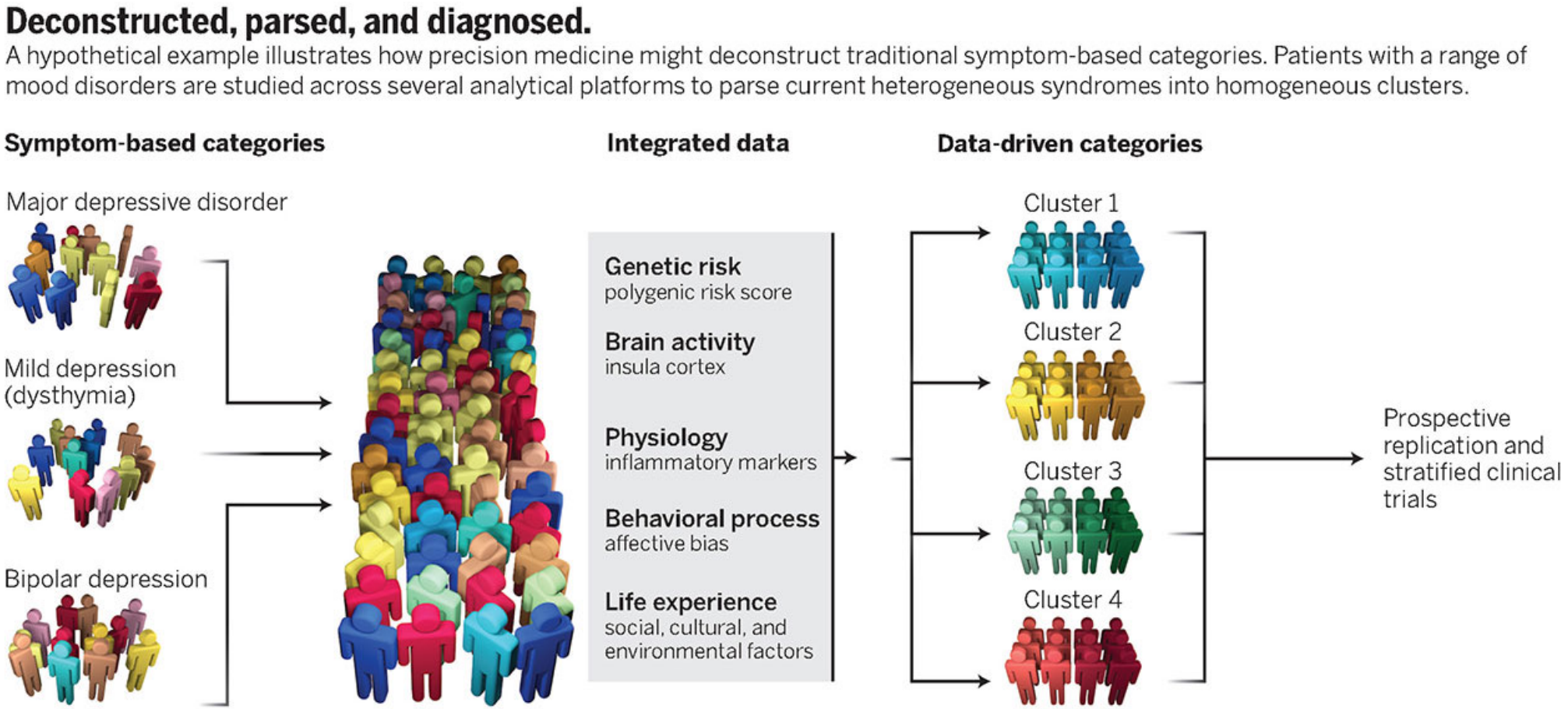
More data means means more interesting and useful science

- Consistently identifying **subtle** fMRI magnitude differences between populations requires a lot of (good quality) data
- Meta-analyses using data can identify things that were ignored in the original studies

If we want to understand brain disease, we need a lot of data

NIMH Research Domain Criteria (RDoC)

We can't study brain disorder X vs controls anymore



Insel & Cuthbert, "Brain disorders? Precisely" *Science* May 1, 2015

Why most people don’t share data

Selfish Reasons

1. It takes time that can be used for other things
2. I collected my data. I’m not giving it away so that other people can publish and take credit for my work

Replication is a nice goal… in theory

Discussion of replicating an experiment using superconducting bar magnets to detect gravity waves in the 1970's (not MRI)

"Should the bar be cast from the same batch of metal? Should we buy the piezoelectric crystals form the same manufacturer as the original ones? Should we glue them to the bar using the same adhesive as before, bought from the same manufacturer? Should the amplifiers be identical in every visible respect, or is an amplifier build to certain input and output specifications "the same" as another amplifier built to the same specifications? Should we be making sure that the length and diameter of every wire is the same? Should we be making sure that the color of the insulation on the wires is the same? Clearly, somewhere one as to stop asking such questions and use a commonsense notion of "the same." the trouble is that in frontier science, tomorrow's common sense is still being formed."

Gravity's Shadow by Harry Collins p 123

Replication is a nice goal… in theory

Discussion of replicating an experiment using superconducting bar magnets to detect gravity waves in the 1970's (not MRI)

> To my astonishment, on reading this passage, Gary Sanders told me that in 1988 or 1989 he had been in a laboratory near Tokyo when a Russian physicist, examining a Japanese group's apparatus, declared their results on Tritum Beta decay to be invalid because they had used wires with red insulation! Apparently the red dye contains traces of radioactive uranium, which can confound the measurements.

and diameter of every wire is the same? **Should we be making sure that the color of the insulation on the wires is the same?** Clearly, somewhere one as to stop asking such questions and use a commonsense notion of "the same." the trouble is that in frontier science, tomorrow's common sense is still being formed."

Gravity's Shadow by Harry Collins p 123

Variation in replications can be good

The most replicated finding in drug abuse research is, “Rats will intravenously self-administer (IVSA) cocaine.”

The types of factors that can reduce your "effect" to a null effect or change the outcome include:

- Catheter diameter or length
- Cocaine dose available in each infusion
- Rate of infusion/concentration of drug
- Age, Sex, Strain, or vendor source of the rats
- Time of day in which rats are run (not just light/dark* either)
- Food restriction status & last food availability
- Pair vs single housing & “Enrichment”
- Experimenter choice of smelly personal care products
- Dirty/clean labcoat (I kid you not)
- Handling of the rats on arrival from vendor
- Fire-alarm
- Cage-change day
- Minor rat illness
- Location of operant box in the room (floor vs ceiling, near door or away)
- Ambient temperature of vivarium or test room
- Schedule- weekends off? seven days a week? 1 hr? 2hr? 6 hr? access sessions
- Schedule- are reinforcer deliveries contingent upon one lever press? five? does the requirement progressively increase with each successive infusion?
- Animal loss from the study for various reasons

We only know this because so many replications were done **and documented**

http://scientopia.org/blogs/drugmonkey/2014/07/08/the-most-replicated-finding-in-drug-abuse-science/

The promise of data sharing?
And the exciting findings are…

- Thanks to shared fMRI big datasets what have we learned about the brain and disease?

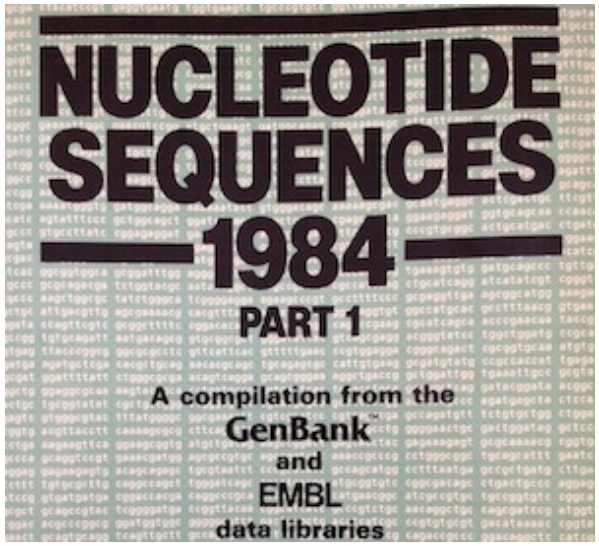
  Seriously, what?

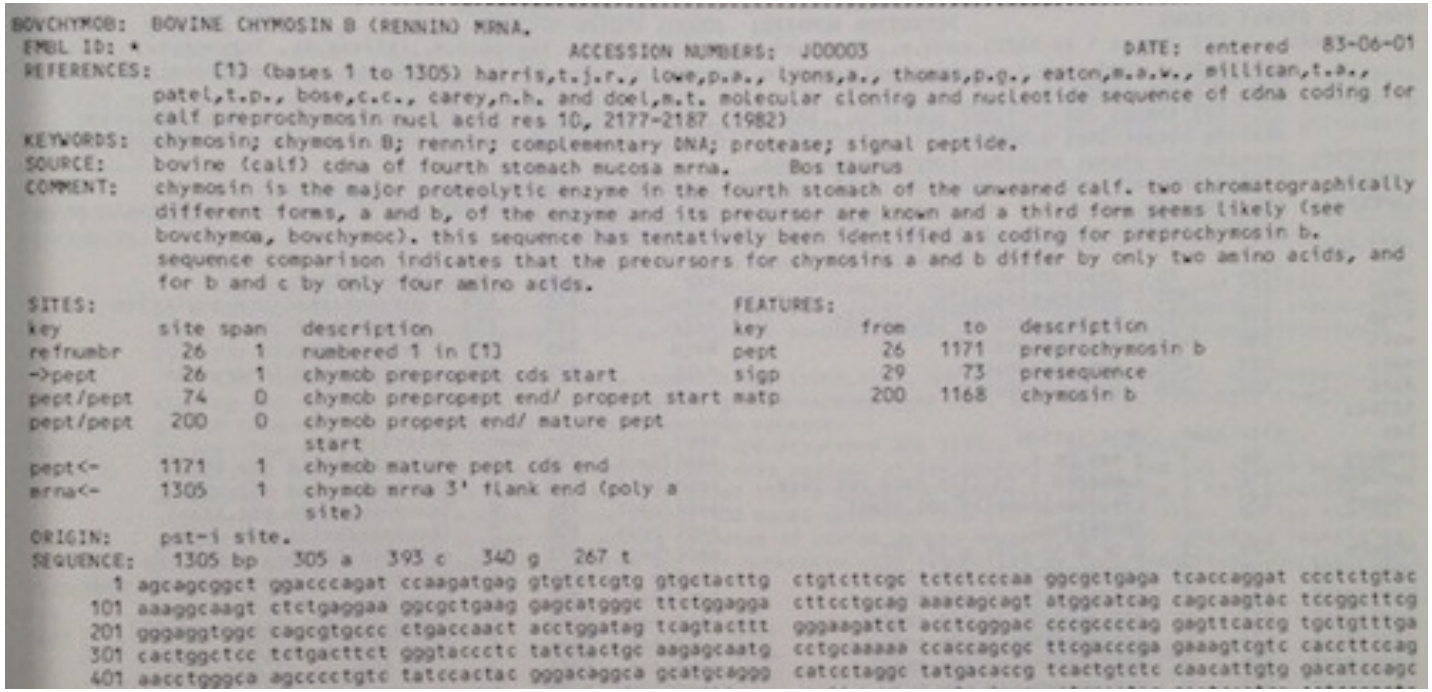
- (There are interesting, but not yet major findings)
  Di Martino, Yan, et al, Molecular Psychiatry 2014
  Laumann, Gordon et al, Neuron 2015

- (A lot of interesting & important stuff, particularly methods development, is happening thanks to these sharing)

- Why is this?
  - The field is still young
  - Sharing data is more than just posting files somewhere
    We need to "harmonize" data so that it can make sense to others

The field is still young
Building big data capabilities takes time





```
BOVCHYMOB:  BOVINE CHYMOSIN B (RENNIN) MRNA.
EMBL ID: *                                  ACCESSION NUMBERS:  J00003                    DATE:  entered  83-06-01
REFERENCES:     [1] (bases 1 to 1305) harris,t.j.r., lowe,p.a., lyons,a., thomas,p.g., eaton,m.a.w., millican,t.a.,
          patel,t.p., bose,c.c., carey,n.h. and doel,m.t. molecular cloning and nucleotide sequence of cdna coding for
          calf preprochymosin nucl acid res 10, 2177-2187 (1982)
KEYWORDS:   chymosin; chymosin B; rennin; complementary DNA; protease; signal peptide.
SOURCE:     bovine (calf) cdna of fourth stomach mucosa mrna.     Bos taurus
COMMENT:    chymosin is the major proteolytic enzyme in the fourth stomach of the unweaned calf. two chromatographically
          different forms, a and b, of the enzyme and its precursor are known and a third form seems likely (see
          bovchymoa, bovchymoc). this sequence has tentatively been identified as coding for preprochymosin b.
          sequence comparison indicates that the precursors for chymosins a and b differ by only two amino acids, and
          for b and c by only four amino acids.

SITES:                                                          FEATURES:
key         site span   description                             key         from     to   description
refnumbr      26    1   numbered 1 in [1]                       pept          26   1171   preprochymosin b
->pept        26    1   chymob prepropept cds start             sigp          29     73   presequence
pept/pept     74    0   chymob prepropept end/ propept start    matp         200   1168   chymosin b
pept/pept    200    0   chymob propept end/ mature pept
                        start
pept<-      1171    1   chymob mature pept cds end
mrna<-      1305    1   chymob mrna 3' flank end (poly a
                        site)
ORIGIN:     pst-i site.
SEQUENCE:   1305 bp   305 a   393 c   340 g   267 t
       1 agcagcggct ggacccagat ccaagatgag gtgtctcgtg gtgctacttg  ctgtcttcgc tctctcccaa ggcgctgaga tcaccaggat ccctctgtac
     101 aaaggcaagt ctctgaggaa ggcgctgaag gagcatgggc ttctggagga  cttcctgcag aaacagcagt atggcatcag cagcaagtac tccggcttcg
     201 gggaggtggc cagcgtgccc ctgaccaact acctggatag tcagtacttt  gggaagatct acctcgggac cccgccccag gagttcaccg tgctgtttga
     301 cactggctcc tctgacttct gggtaccctc tatctactgc aagagcaatg  cctgcaaaaa ccaccagcgc ttcgacccga gaaagtcgtc caccttccag
     401 aacctgggca agccccctgt catccactac gggacaggca gcatgcaggg  catcctaggc tatgacaccg tcactgtctc caacattgtg gacatccagc
```

- USA started funding the public GeneBank in 1982
- The National Center for Biotechnology Information at NLM launched & look over GeneBank in 1989
- Basic Local Alignment Search Tool (BLAST) algorithm to rapidly match sequences came out in 1990
- The first human genomes were released in 2003.
- 23andMe started commercial genome sequencing & sharing in 2007
- Nearly real time tracking of dangerous bacteria at NIH Clinical center in 2012 (Snitken, Sci Transl Med 2012)

Images from: http://www.davelunt.net/evophylo/2012/12/printing-out-genbank-nucleotide-sequences-1984/

General challenges

- Universal big data issues
  - Giving multiple people quick access to data
    (The amount of data was an issue, but, except for unreconstructed fMRI data, fMRI is smaller than a lot standard big data applications)
  - Security: Different types of data need different levels of protection
- Application specific issues
  - Quality control with less human interaction
    - If it's enough data to be interesting, it's too much to have each group examine everything interactively
  - Meaningful queries of data
  - Documentation and provenance

Almost this whole talk could be generic data sharing, but the type of data matters

The Organization for Human Brain Mapping created the Committee on Best practices in Data Analysis & Sharing (COBIDAS)

All they could agree on was what to report, not what to do

Aspect	**Notes**	**R = Required r = Recommended**
Number of subjects	Elaborate each by group if have more than one group.	
Subjects approached		r
Subjects consented		r
Subjects refused to participate	Provide reasons.	r
Subjects excluded	If any; provide reasons.	R
Subjects participated	Final number of subjects included in the statistical analysis, specifying if that number varies between different anlayses.	R
Inclusion Criteria and Descriptive Statistics	Elaborate each by group if have more than one group.	
Age	Mean, standard deviation and range.	R
Gender	Absolute or relative frequencies	R
Race & Ethnicity	Per guidelines of NIH or other relevant agency	R
SES, Education	Specify measurement instrument used; may be parental SES and education if study has minors.	r
IQ	Specify measurement instrument used.	r
Handedness	Absolute or relative frequencies; basis of handedness-attribution (self-report, EHI, other tests)	R
Exclusion criteria	Describe any screening criteria, including those applied to “normal” sample such as MRI exclusion criteria.	R

From Presentation by Tom Nichols
NIMH Workshop: Harmonize This!
6/19/2015

What to store/report?

- MRI data acquisition information
  - Basic scanning parameters vs detailed pulse sequence descriptions?
  - The strength/brand of the scanner vs full hardware specifications?
  - Inglis “A checklist for fMRI acquisition methods reporting in the literature” *The Winnower* 2015 DOI:10.15200/winn.143191.17127
- Other information:
  - Task designs vs presentation scripts?
  - Behavioral and physiological data and keys to understand what it means
  - Statistical maps vs reconstructed vs raw fMRI data?
  - Processing steps (Neurimaging Data Model NDM) vs code
  - Carp “The secret lives of experiments: Methods reporting in the fMRI Literature” *NeuroImage* 2012

The failure of the FMRI Data Center

Started in 1999

- Scientists weren't motivated to contribute
  - No social pressure
  - It was really hard to enter data and information and submit it
- fMRIDC needed to organize data in whatever format it got (unorganized hard drives people pulled from their computers)
- Removing identifiable information
- Data shared by mail
- Short-term funding disappeared

Van Horn, J.D., Gazzaniga, M.S., 2012.
Why share data? Lessons learned from the fMRIDC. NeuroImage

Ethics of data sharing

- Internal review boards require every study to have a purpose. Data can't be used outside the scope of its purpose of collection
  - It can be hard to share data collected without sharing mentioned in the protocols
- If data sharing is clinically useful, we will learn unexpected things about individuals' future health
- Anonymity is impossible to maintain by computer security alone

Anonymity doesn't exist with big data

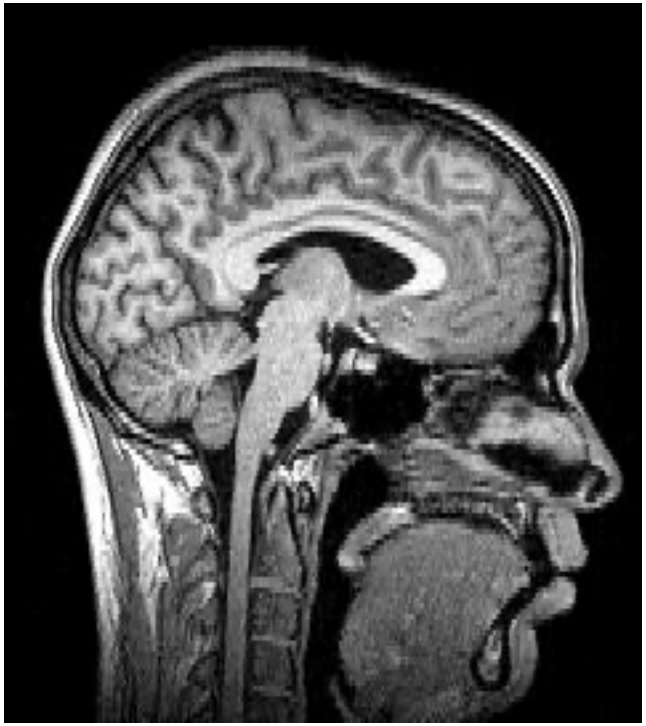
Sisters discovered they didn't share a father

A neuroscientist who was adopted used 23andMe and Facebook to identify some probable half-siblings (but decided not to contact them)

23 and You, by Virginia Hughes (12/4/2013)

https://medium.com/matter-archive/23-and-you-66e87553d22c

MRI and anonymity



Facebook post:
"Hey look at this cool picture of my brain I got as thanks for participating in a study"

The data is publicly shared. Someone takes the picture, matches it to a slice in shared data and then gets the genome & diagnostic info linked to the study.

Brain slice from: http://en.wikipedia.org/wiki/Neuroimaging#mediaviewer/File:Sagittal_brain_MRI.JPG

How to prevent loss of privacy?

- MRI brain slices and genomes are not considered identifiable information in US health privacy law
  (We talk about de-identification, not anonymization)
- Try to make this type of matching difficult
  - Blur faces on high quality images
  - Remove easily traceable information
  - Don't give participants the results (unethical?)
- Lock data behind data use agreements
  (makes sharing more difficult)

Why fMRI data sharing is improving

- More examples means the hardest work is already done
  - We have examples on how to organize data
  - We have quality control & data analysis examples
  - We have examples of language for protocols and consent forms
  - We have data use agreements
  - Sharing big data is now common
  - The general technology needed to securely store, share, run computations, and visualize big fMRI data exists
- There are big investments in data sharing
  - NIH repositories
  - General data sharing rules for grantees
  - Grants to projects with a data sharing focus

Resting State fMRI has some of the early logistical successes

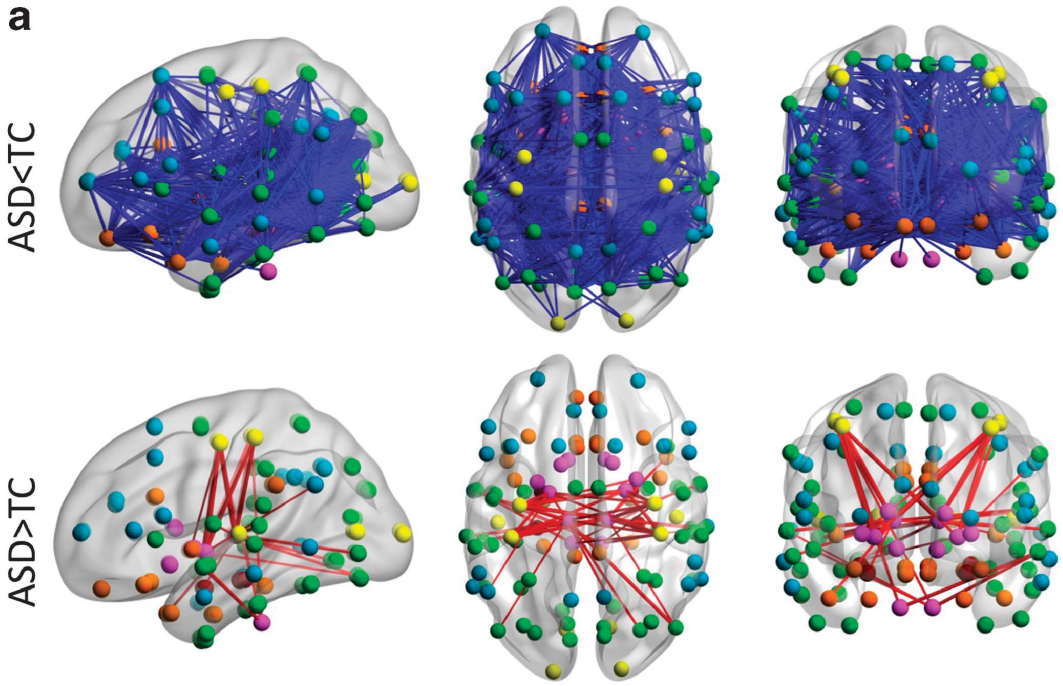
- Bypasses potential sources of variation associated with task probes
- Commonly included as an add-on in task activation studies
  - decreases perceived value
  - Increases willingness to share
- Striking similarity in networks observed across laboratories
- Problems: More variation than originally assumed
- Huge successes are methodological
  - Automated quality assessment
  - Better understanding of sources of variability
  - Comparisons of data analysis tools on the same data sets

Text copied or heavily based on presentation by Michal Milham
NIMH Workshop: Harmonize This! 6/19/2015

The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism

Di Martino et al, Molecular Psychiatry 2014

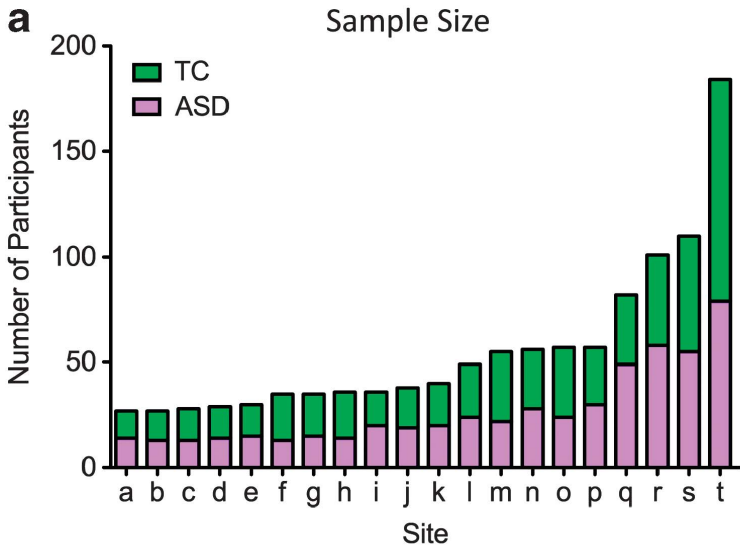
a



b

Functional Divisions	Primary SM		Unimodal		Heteromodal		Paralimbic		Limbic		Subcortical	
ASD vs.TC	*Hypo*	*Hyper*	*Hypo*	*Hyper*	*Hypo*	*Hyper*	*Hypo*	*Hyper*	*Hypo*	*Hyper*	*Hypo*	*Hyper*
	n (%)		*n (%)*		*n (%)*		*n (%)*		*n (%)*		*n (%)*	
Primary SM	6 (13%)											
Unimodal	88 (20%)		276 (29%)									
Heteromodal	32 (13%)		166 (16%)		57 (21%)							
Paralimbic	34 (19%)		147 (19%)		104 (24%)		57 (37%)					
Limbic	7 (18%)		14 (8%)		7 (7%)		14 (19%)		1 (17%)			
Subcortical		20 (17%)		18 (3%)		1 (0%)	10 (5%)	1 (0%)		1 (2%)	2 (3%)	
Total n (%)	**167 (16%)**	**20 (2%)**	**691 (18%)**	**18 (0%)**	**366 (15%)**	**1 (0%)**	**366 (20%)**	**1 (0%)**	**43 (11%)**	**1 (0%)**	**12 (1%)**	**41 (3%)**

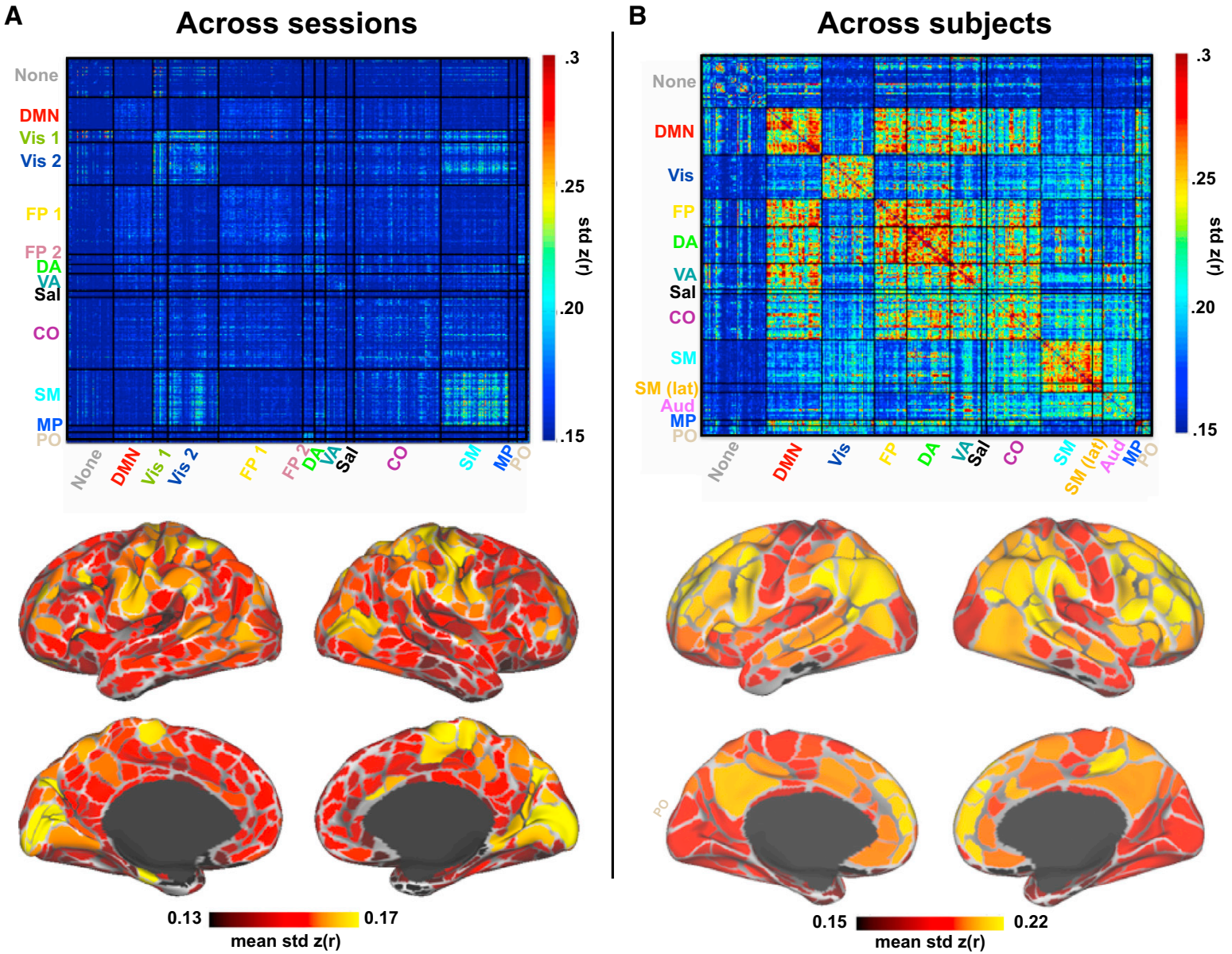
a



Functional System and Areal Organization of a Highly Sampled Individual Human Brain

Laumann, et al, *Neuron* 2015

http://myconnectome.org/ + 120 subjects from Washington University



Mostly resting state retrospective data sharing

	Contributors	Release Date	Sample Info	Phenotypic Data/Imaging Data
	ADHD-200 Jan Buitelaar, F. Xavier Castellanos, Damien Fair, Bea Luna, Michael Milham, Stewart Mostofsky, Joel Nigg, Steve Petersen, Brad Schlaggar, Julie Schweitzer, Katerina Velanova, Yu-Feng Wang, Yu-Feng Zang	**Released**	383 ADHD 491 598 (ages 7-21)	DSM-IV diagnostic status, dimensional ADHD symptom measures, age, sex, intelligence quotient and lifetime medication status *R-fMRI, MPRAGE*
	Beijing Enhanced Yu-Feng Zang	**Released**	Community Sample n=180	FSIQ, VIQ, PIQ (n=55) *R-fMRI, DTI, MPRAGE (n=180)*
	North Shore - LIJ Ashesh Mehta, Stephan Bickel, Laszlo Entz	**Released**	Surgical Epilepsy Patients n=6	Age, sex, lesion location *R-fMRI, MPRAGE*
	NYU Institute for Pediatric Neuroscience - Cocaine Michael Milham, Adriana Di Martino, Clare Kelly, Maarten Mennes, F. Xavier Castellanos	**Released**	Cocaine dependent n=29 Healthy controls n=24	Psychiatric assessment results, symptom severity measures, trait questionnaires *R-fMRI*
	Cleveland CCF Mark Lowe, Erik Beall, Michael Phillips	**Released**	Community Sample n=31	Phenotypic Data - age, gender Heart rate and breathing obtained during R-fMRI *R-fMRI, MPRAGE*
	TRAIN-39 Art Kramer, Michelle Voss, Kirk Erickson, Ruchika Prakash	**Released**	39 young adults trained 20 hours each on a complex video game	Phenotypic Data - age, gender, performance (learning data) on the video game
	Power 2012 Power, J.D., K.A. Barnes, A.Z. Snyder, B.L. Schlaggar, and S.E. Petersen	**Released**	77 children, adolescents, and adult controls	Phenotypic Data - age, gender
	Adelstein 2011 Adelstein, J.S., Shehzad, Z., Mennes, M., DeYoung, C.G., Zuo, X.N., Kelly, C., Margulies, D.S., Bloomfield, A., Gray, J.R., Castellanos, F.X., Milham, M.P.	**Released**	39 healthy adults, 18 male	Behavioral assessments and phenotypic information
	ABIDE Autism Brain Imaging Data Exchange	**Released**	539 individuals with ASD, 573 typical controls	Phenotypic and diagnostic information *R-fMRI, MPRAGE*
	COBRE Center for Biomedical Research Excellence	**Released**	72 patients with Schizophrenia, 75 healthy controls	Phenotypic and diagnostic information *R-fMRI, MPRAGE*
	CoRR Consortium for Reliability and Reproducibility	**Released**	1630 Subjects	Phenotypic and diagnostic information *R-fMRI, MPRAGE, DTI, ASL*
	HNU Short TR Short-TR Eyes-open/Eyes-closed Resting State fMRI Data	**Released**	45 Subjects	Phenotypic and diagnostic information *R-fMRI, T1*
	ARRA NIDA Multimodal Imaging Treatment of ADHD Followup Neuroimaging Study	**Pending**	129 Subjects	Phenotypic and diagnostic information *R-fMRI, T1*

http://fcon_1000.projects.nitrc.org/indi/IndiRetro.html

Mostly resting state prospective data sharing

	Contributors	Release Date & Frequency	Sample Info	Phenotypic Data/Imaging Data
	Beijing: Eyes Open Eyes Closed Study Yu-Feng Zang	**Completed** n=48	Community Sample	3 resting state scans: first eyes closed, second and third counterbalanced for eyes open or closed *R-fMRI, DTI, MPRAGE*
	Beijing: Short TR Study Yu-Feng Zang	**Completed** n=28	Community Sample	Various dimensional psychiatric scales & behavioral performance measures R-fMRI scans with 2 second TRs as well as R-fMRI scans with 0.4 second TRs *R-fMRI, DTI, MPRAGE*
	NKI/Rockland Sample Bharat Biswal, F. Xavier Castellanos, Barbara Coffey, Stan Colcombe, David Guilfoyle, Matthew Hoptman, Dan Javitt, Harold S. Koplewicz, Bennett Leventhal, Larry Maayan, Maarten Mennes, Michael Milham, Kate Nooner, Nunzio Pomara	**Completed** n=207	Psychiatrically-Evaluated Sample (ages 4-90)	Intelligence testing, psychiatric diagnostic interview, executive function performance measures, dimensional psychiatric scales and laboratory results *R-fMRI, DTI, MPRAGE*
	NYU Institute for Pediatric Neuroscience Sample Michael Milham, Adriana Di Martino, Clare Kelly, Maarten Mennes, F. Xavier Castellanos	**Ongoing** 25-50 Quarterly n=49	Psychiatrically-Evaluated Sample (ages 6-55)	FSIQ, VIQ, PIQ *R-fMRI, DTI, MPRAGE*
VTC	**Virginia Tech Carilion Research Institute** Cameron Craddock, Stephen LaConte, The Neuro Bureau	**Ongoing** 25+ Quarterly n=25	Community Sample	Test-Retest R-fMRI scans *R-fMRI, MPRAGE*
	The Neuro Bureau/Berlin Mind and Brain Institute Daniel Margulies, Arno Villringer, The Neuro Bureau	**Ongoing** 25+ Quarterly n=50	Community Sample	Morophometric information and affective trait scales *R-fMRI, MPRAGE*
	The Quiron-Valencia Sample Luis Marti-Bonmati, Maria de la Iglesia Vaya, the Spanish Resting State Network	**Ongoing** 25+ Quarterly n=45	Community Sample	Demographic information *R-fMRI, MPRAGE*
	NKI-RS Multiband Imaging Test-Retest Pilot Dataset	**Pilot Sample** n=24	Community Sample (scanned two times; 1 week apart)	Demographic information *multiband R-fMRI, multiband DTI, breath hold scan, eye movement calibration scan, visual stimulation scan*
	NKI-RS Enhanced Sample	**Ongoing** n=195	Community Representative Sample	Demographic information *multiband R-fMRI, multiband DTI, breath hold scan, eye movement calibration scan, visual stimulation scan*
ImageNotAvailable	**Huaxi MR Research Center (HMRRC),West China Hospital of Sichuan University** Kaiming Li	30 participants	Chess Grandmasters	rfMRI, sMRI, DTI, MetaData
Coming Soon	**Brain Imaging Research Center at UAMS** Andrew James, Clint Kilts	25+ Quarterly	Psychiatrically Evaluated Sample (Cocaine dependent); Community Sample	Psychiatric diagnostic interview, visuospatial, language fluency, executive function
Coming Soon	**The Mind Research Network** Vince Calhoun, Juan Bustillo, Andy Mayer, Rex Jung	30-50/year	Psychiatrically screened sample	Various dimensional psychiatric scales & behavioral performance measures
Coming Soon	**Duke University Medical Center** David Madden, Nan-kuei Chen	10+ Quarterly	Community-dwelling adults	Various behavioral performance measures
Coming Soon	**Harvard-MGH** Randy Buckner	50+ Quarterly	Community Sample	Extended demographic information and a mix of trait/performance measures, as well as personality assessments.
Coming Soon	**Kennedy Krieger Institute** Stewart Mostofsky	15-20 Quarterly	Psychiatrically Evaluated Sample (ages 8-12)	Demographic information, psychiatric diagnostic interview, psychiatric/ behavioral questionnaires, cognitive testing (IQ/achievement), executive function measures, motor function measures
Coming Soon	**University of Texas at Austin/UCLA** Russ Poldrack, Eliza Congdon	15-20 quarterly	Community sample screened for psychiatric and medical conditions	Various psychological scales and behavioral performance measures (some of which will be embargoed). Whole-transcriptome peripheral gene expression (to be embargoed).
Coming Soon	**MPI-Leipzig** Daniel Margulies, Arno Villringer, The Neuro Bureau	25+ Quarterly	Community Sample	Psychological questionnaires (e.g., PANAS, PDI, DSQ)
Coming Soon	**Stanford University** Michael Greicius	60 participants	Alzheimer's disease, non-AD dementias, older controls	MMSE, CDR, neuropsychological measures

http://fcon_1000.projects.nitrc.org/indi/IndiPro.html

Examples of cross-site data collection

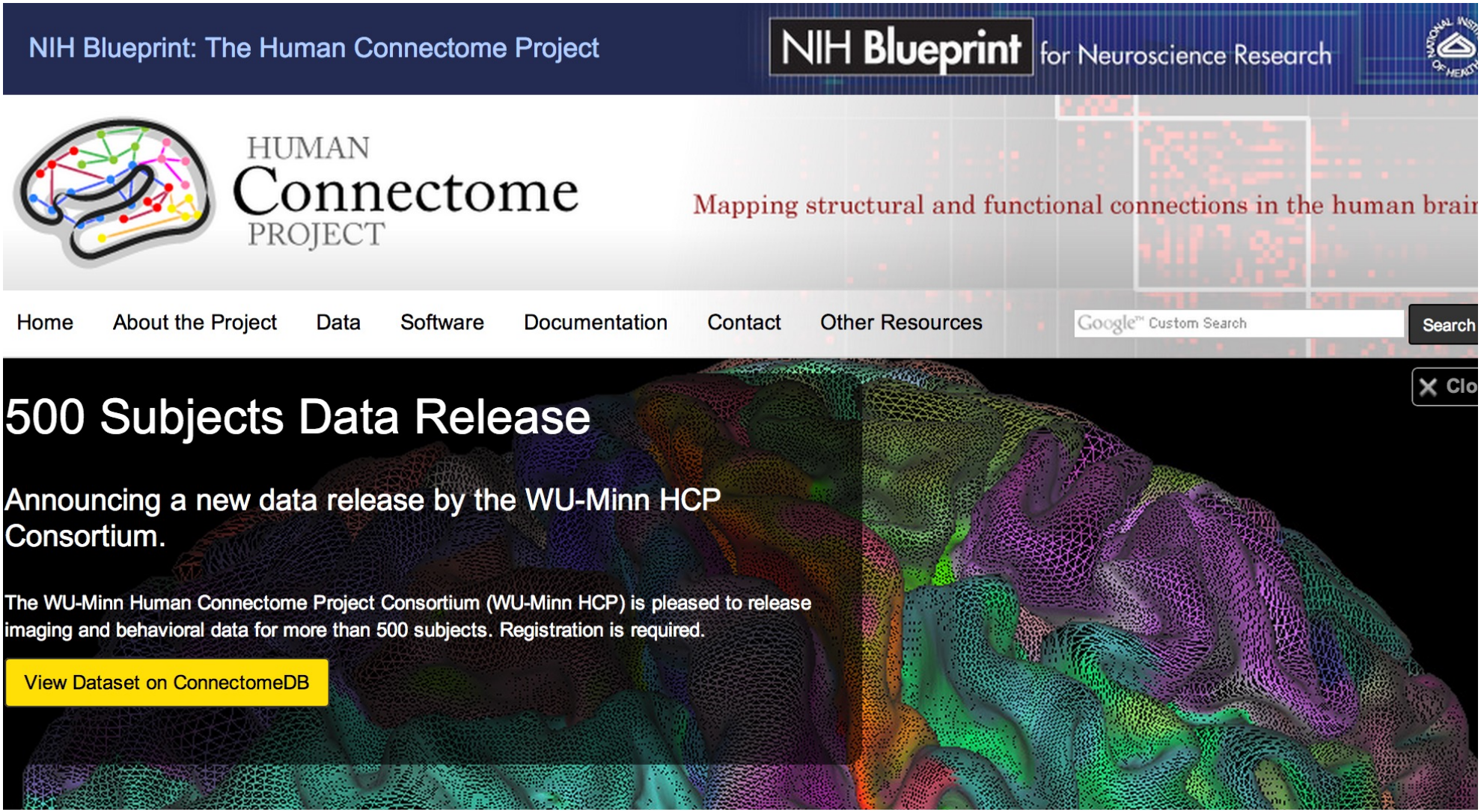
- Consortia with common experimental paradigms
  - ADNI (Alzheimer's Disease)
  - Human Connectome Projects
- Consortia with common populations
  - NDAR (National Database for Autism Research)
  - FITBAR (Traumatic Brain Injury)
  - dbGaP (Genotypes + phenotypes)
  - ABIDE (Autism)
  - ADHD-200
  - ENIGMA (Imaging + Genetics)
  - FBIRN (Standardization across sites)

More places to put or access data

- Full fMRI time series
  - Openfmri.org
  - central.xnat.org
  - fcon_1000.projects.nitrc.org (and nitrc.org)
  - incf.org/resources/data-space/
  - COINS coins.mrn.org
  - www.painrepository.org
  - nidb.sourceforge.net
  - ida.loni.usc.edu
  - Alzheimers: gaain.org
  - Pediatrics: pingstudy.ucsd.edu
- Statistical Maps or clusters
  - BrainMap.org
  - Neurosynth.org
  - Neurovault.org
  - anima.modelgui.org Meta-analyses
- Atlases & templates!

Large investments in collecting data for sharing

Centralized data collection and preliminary analyses



While important, these projects are rare, risky, and expensive

Human Connectome Project example

How do you share 50+ Terabytes of fMRI data?

- If you want the first 500 subjects of data, they'll mail you five 4 Terabyte drives for $750.
- Parts of the data are available online to download or explore.

National Database for Autism Research

- NDAR data lives on the aws.amazon.com
- Data is at one location
- Users need accounts & access, but don't need local data
- Computer use costs money, but you can scale with the computational power you need & you don't need to buy super-powerful local computers
- As fast or slow as other AWS websites

Documentation and provenance examples

- Brain Imaging Data Structure (BIDS) bids.neuroimaging.io
- Neuroimaging Data Model nidm.nidash.org/
  - Automated ontology for understanding statistical maps
- NIF (Neuroscience Information Framework)
  - Ontology for all of neuroscience
- Data processing scripts
- AFNI software realtime output and history flags

Washington University - University of Minnesota Consortium of the Human Connectome Project: Data Use Agreement

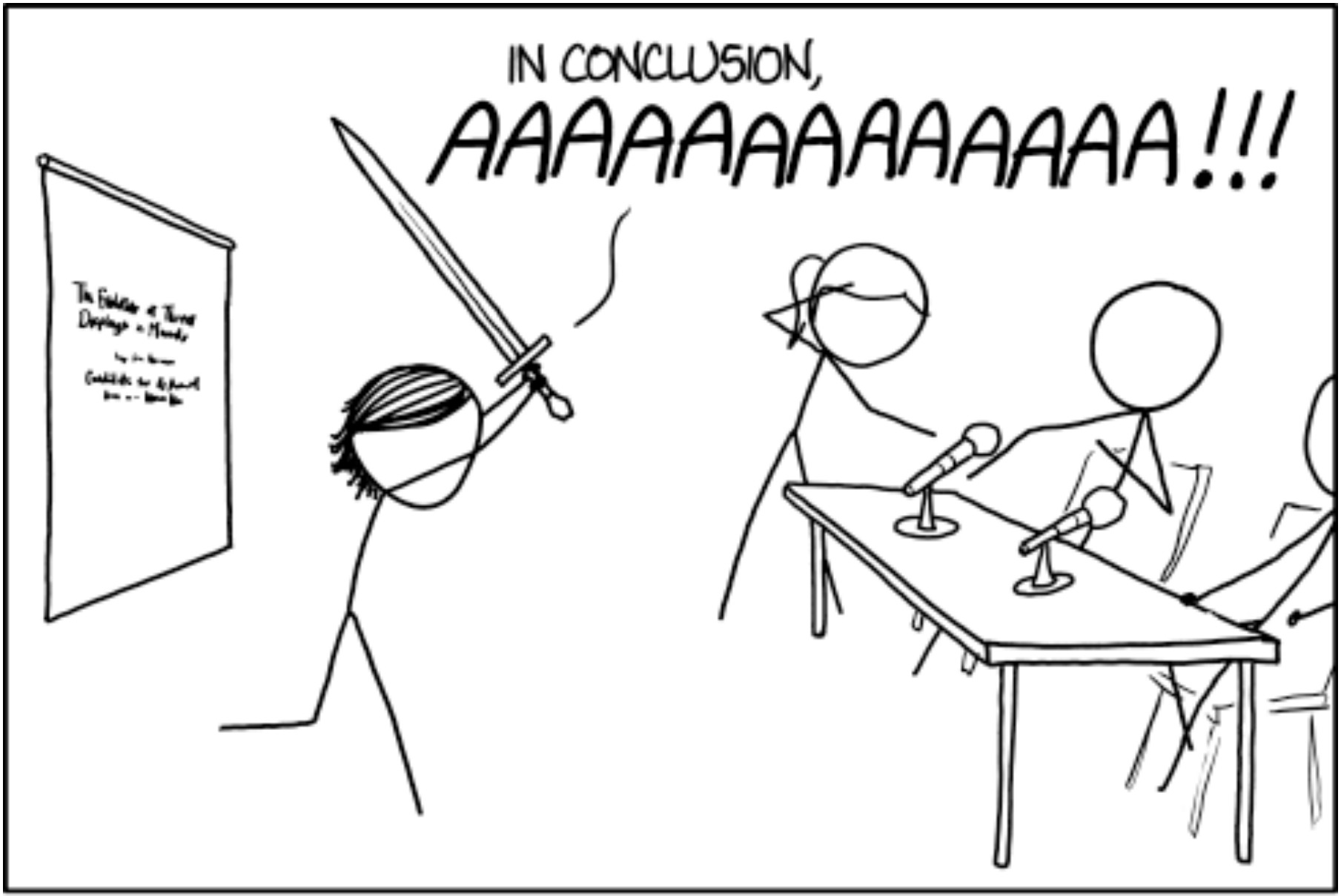
(abridged)

- I will not attempt to establish the identity of or attempt to contact any of the included human subjects.
- I understand that under no circumstances will the code that would link these data to Protected Health Information be given to me, nor will any additional information about individual human subjects be released to me.
- I will comply with all relevant rules and regulations imposed by my institution. This may mean that I need my research to be approved or declared exempt by a committee that oversees research on human subjects, e.g. my IRB or Ethics Committee. The released HCP data are not considered de-identified, insofar as certain combinations of HCP Restricted Data (available through a separate process) might allow identification of individuals.
- I will acknowledge the use of WU-Minn HCP data… when publicly presenting any results or algorithms that benefitted from their use.
- Authors of publications or presentations using WU-Minn HCP data should cite relevant publications describing the methods used by the HCP…
- Failure to abide by these guidelines will result in termination of my privileges to access WU-Minn HCP data.

Conclusions

- fMRI data sharing is growing very rapidly
- It has the potential to greatly increase our understanding of the brain
- Practical, technological, financial, and ethical barriers remain, but are disappearing fast!

Questions?



http://xkcd.com/1403/