Best practice in data-sharing in neuroimaging… and more

Thomas Nichols

University of Warwick

http://warwick.ac.uk/tenichols

OHBM Committee on Best practice in Data Analysis & Sharing (COBIDAS) History

- Created at OHBM 2014
  - Prompted by "OHBM Council Statement on Neuroimaging Research and Data Integrity"
- Charged with
  1. Identifying best practices of data analysis and data sharing in the brain mapping community
  2. Preparing a white paper organising and describing these practices
  3. Seeking input from the OHBM community
  4. Publishing these recommendations

COBIDAS Status

- Membership

Simon Eickhoff	Alan Evans	Michael Hanke	Nikos Kriegeskorte
Michale Milham	Russel Poldrack	Jean-Baptiste Poline	Erika Proal
Bertrand Thirion	David van Essen	Tonya White	BT Thomas Yeo
Thomas Nichols			

- Will not prescribe practice (mostly)
- Rather focus on what to report
  - To support open and reproducible research

COBIDAS Status

- fMRI
  - Task & rest
- Divisions of fMRI Practice
  - Experimental design reporting
  - Image acquisition reporting
  - Preprocessing reporting
  - Statistical modeling
  - Results reporting
  - Data sharing
  - Replication and reproducibility
- Produce white paper
  - To be commented on, and ultimately approved by OHBM members

COBIDAS Document Form

- For each division
  - Principals of open and reproducible research
  - When feasible, recommendations for practice
  - Detailed tabular listing of what to report

Best Practices in Data Analysis and Sharing in Neuroimaging using MRI

A report by
Committee on Best Practices in Data Analysis and Sharing (COBIDAS)
Organization for Human Brain Mapping (OHBM)
June X, 2015

A Draft For Consideration by The Membership of OHBM

0. Introduction

The Organization for Human Brain Mapping (OHBM) Committee on Best Practices in Data Analysis and Sharing (COBIDAS) was created by OBHM Council in June 2014 on the basis of the "OHBM Council Statement on Neuroimaging Research and Data Integrity" [link]. COBIDAS was charged with (i) identifying best practices of data analysis and data sharing in the brain mapping community, (ii) preparing a white paper organising and describing these practices, and (iii) seeking input from the OHBM community before ultimately (iv) publishing these recommendations. Dr. Nichols was named as chair and invited nominations from the OHBM membership in July 2015. From over 100 nominees Dr. Nichols selected a dozen experts from the membership that covered reflected the diversity of OHBM, with the final list approved by Council. The different constituencies considered included: Researchers focusing in cognitive applications, clinical applications, methods and database developers; different geographic areas; gender; representation of junior researchers; and, to facilitate communication within OHBM leadership, at least one member from Council and one member from the OHBM Program Committee. The panel of 13 members (including chair) met on a regular basis from October 2014 through May 2015, releasing a draft of this document to the full OHBM membership for comment in June 2015. After a comment period, during which input was solicited from members, a revised document was created and distributed to the membership for an up/down vote. In July XX 2015, an e-ballot was held which approved the document with a simple majority vote. It should be noted that while best practice white papers like this are not uncommon (see, e.g., [Alsop2014,Kanal2013,Gilmore2013]), they are generally authored by and represent the consensus of a small committee or at most a special-interest section of a larger professional body. Hence we are excited to present this work with the explicit consent of a plurality of the OHBM membership.

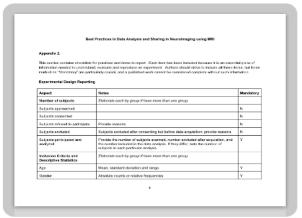
Experimental Design Reporting

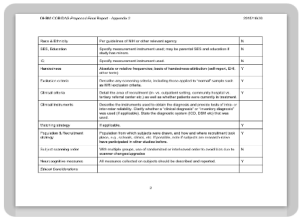
Aspect	**Notes**	**R = Required r = Recommended**
Number of subjects	Elaborate each by group if have more than one group.	
Subjects approached		r
Subjects consented		r
Subjects refused to participate	Provide reasons.	r
Subjects excluded	If any; provide reasons.	R
Subjects participated	Final number of 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

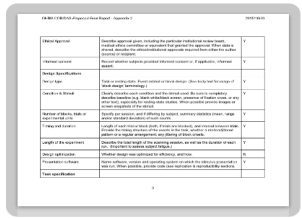
Example: Experimental Design Reporting

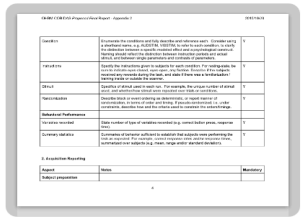
Aspect	Notes	Mandatory
Number of subjects	*Elaborate each by group if have more than one group.*	
Subjects approached		N
Subjects consented		N
Subjects refused to participate	Provide reasons.	N
Subjects excluded	Subjects excluded after consenting but before data acquisition; provide reasons.	N
Subjects participated and analyzed	Provide the number of subjects scanned, number excluded after acquisition, and the number included in the data analysis. If they differ, note the number of subjects in each particular analysis.	Y
Inclusion Criteria and Descriptive Statistics	*Elaborate each by group if have more than one group.*	
Age	Mean, standard deviation and range.	Y
Gender	Absolute counts or relative frequencies	Y
Race & Ethnicity	Per guidelines of NIH or other relevant agency	N
SES, Education	Specify measurement instrument used; may be parental SES and education if study has minors.	N
IQ	Specify measurement instrument used.	N
Handedness	Absolute or relative frequencies; basis of handedness-attribution (self-report, EHI, other tests)	Y
Exclusion criteria	Describe any screening criteria, including those applied to “normal” sample such as MRI exclusion criteria.	Y
Clinical criteria	Detail the area of recruitment (in- vs. outpatient setting, community hospital vs. tertiary referral center etc.) as well as whether patients were currently in treatment.	Y

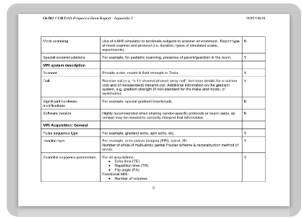
COBIDAS: 100+ reporting items

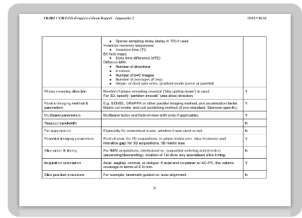

1

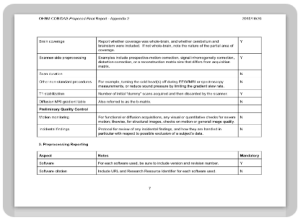

2

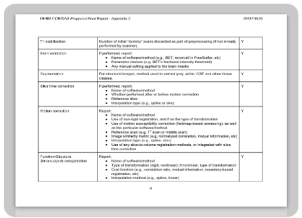

3

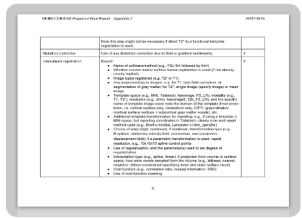

4

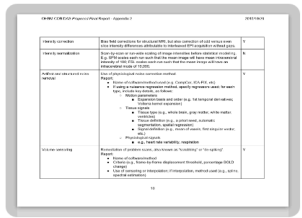

5

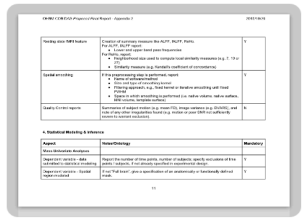

6

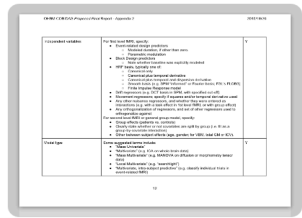

7

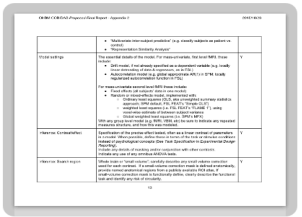

8

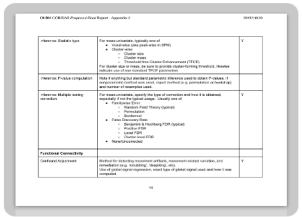

9

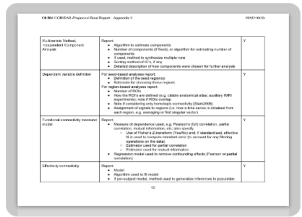

10

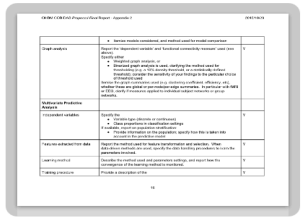

11

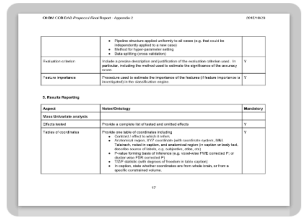

12

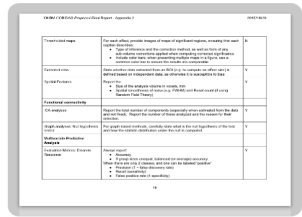

13

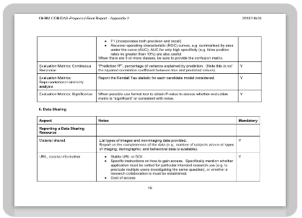

14

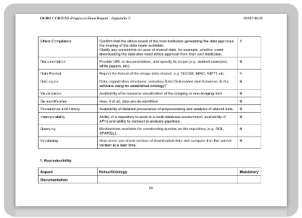

15

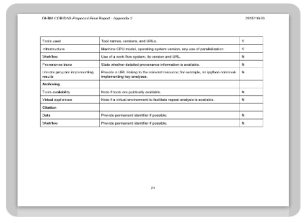

16


17


18


19


20


21

Scientific Data's Meta Data Standard

Subject	organism	organism part	developmental stage	age	sex	laterality	Additional columns…
Patient1	Homo sapiens	brain	adult	43	male	Right	…
Patient2	Homo sapiens	brain	child	37	male	Left	…
Patient3	Homo sapiens	brain	adult	23	female	Right	…
…	…	…	…	…	…	…	…

Subject	Scan Name	Scanning Method	Instrument name	Instrument manufacturer	magnetic field strength value	magnetic field strength unit	coil type	number of coils	Type of MRI assay
Patient1	Scan1	Magnetic Resonance Imaging	Siemens 3T TIM Trio	Siemens	3	tesla	birdcage	12	T1-weighted image
Patient2	Scan2	Magnetic Resonance Imaging	General Electric MR750 3 T scanner	General Electric	3	tesla	head	8	Multi-Echo FLASH
Patient3	Scan3	Magnetic Resonance Imaging	Magnetom	Siemens	7	tesla	head	8	BOLD image
…		…	…	…	…	…	…	…	

Voxel measurements									
field of view	field of view unit	matrix size	matrix size unit	resolution value	resolution unit	repetition time	repetition time unit	echo time	echo time unit
		256x256x192	millimeter	1x1x1	millimeter	2170	millisecond	4.33	millisecond
		64x64x24	millimeter	3.125x3.125x6	millimeter	4000	millisecond	12	millisecond
	27 square centimeter	256 x 256	millimeter			7.3	millisecond	3	millisecond
…	…	…	…	…	…	…	…	…	…

Raw Data File	Data Repository	Data Record Accession	Data Processing Method	Derived Data File	Derived Data Repository	Derived Data Record Accession
data file name	data repository name	accession or doi		derived data file name	data repository name	accession or doi
data file name	data repository name	accession or doi				
data file name	data repository name	accession or doi				
…	…	…				

ISA-Tab: 'Investigation' (the project context), 'Study' (a unit of research) and 'Assay' (analytical measurement) in Tabular format

Replicability, Reproducibility, Repeatability

- COBIDAS report
  - Peng's "Reproducibly": Same data, same code, different researchers
  - Replication: Different data, different researchers, different methods… same conclusion
- ISO *repeatability* (ISO 3534-2:2006 3.3.5) *same-site test-retest*
  - Same "method", "test or measuring facility", same "operator" & "equipment" on "identical test/measurement items" "within short intervals of time".
- ISO *reproducibility* (ISO 3534-2:2006 3.3.10) *between-site test-retest*
  - Same "method" on "identical test/measurement items" in **different** "test or measurement facilities", **different** "operators using different equipment"
- ISO terms adopted by…
  - Radiological Society of North America (RSNA) Quantitative Imaging in Biomarkers Alliance (QIBA) group (https://www.rsna.org/QIBA/) published Kessler at al. (Stat Methods Med Res, 2014)

COBIAS Status

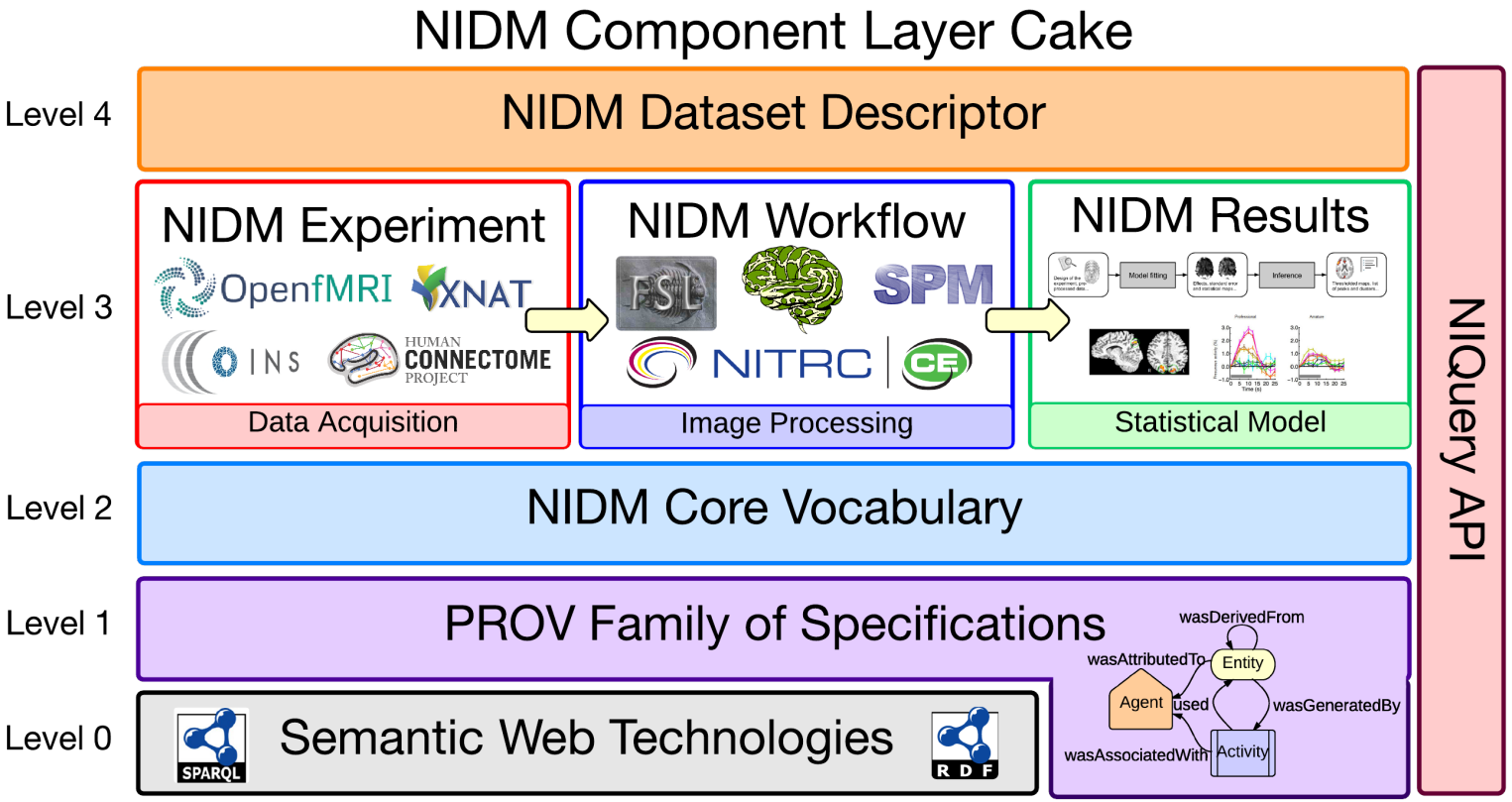
- 20th October
  - Manuscript posted on OHBM website
  - Open for comment for 4 weeks (until 17th Nov.)
- End of November
  - Member comments integrated
  - Final draft, approved by Council
- December
  - Finalized manuscript posted for up/down vote
  - Upon approval, submission for publication

Some Useful Tools & Projects

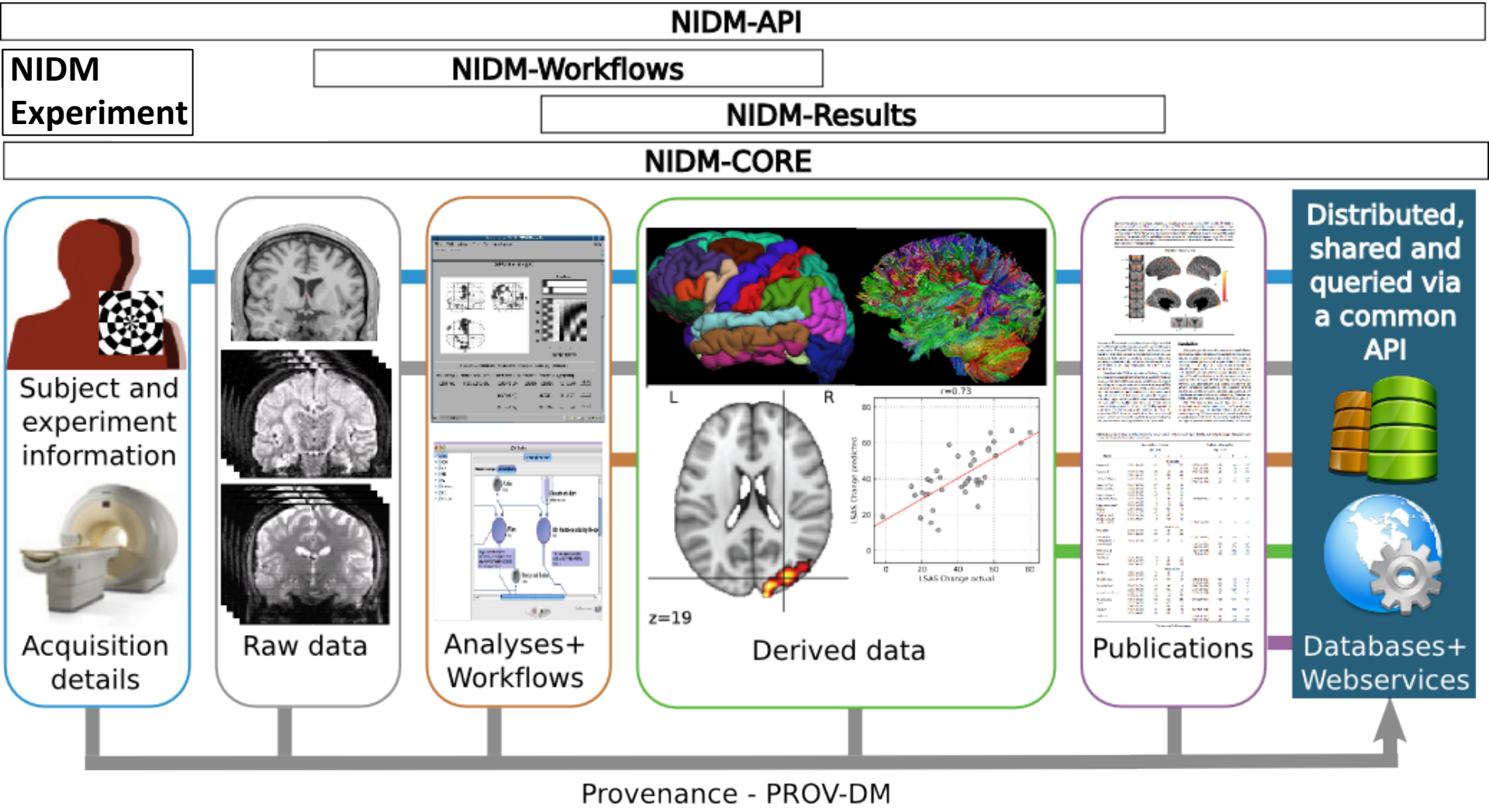
- NIDM
- Neurosynth & Reverse Inference
- NeuroVault

INCF's Neuroimaging Data Model (NIDM)

- Collaborative effort to represent all aspects of neuroimaging experiments, data collection & analysis in semantic web, machine-readable form

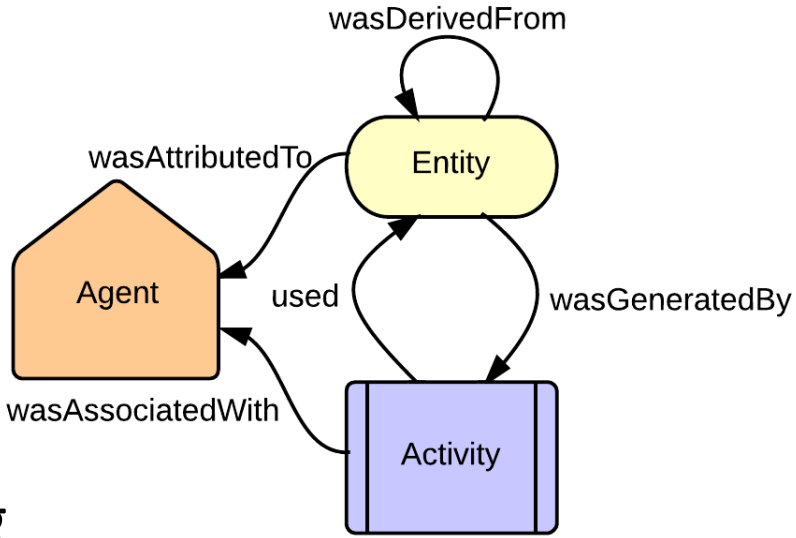


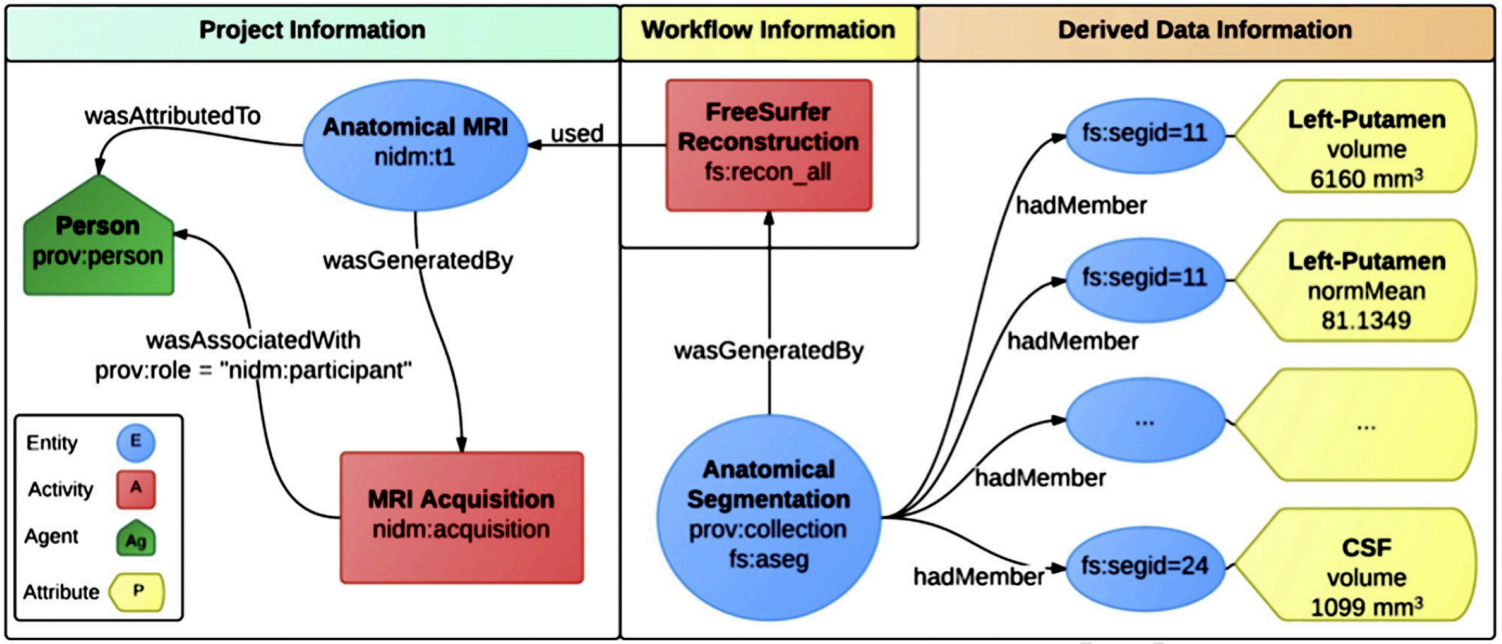
NIDM Overview



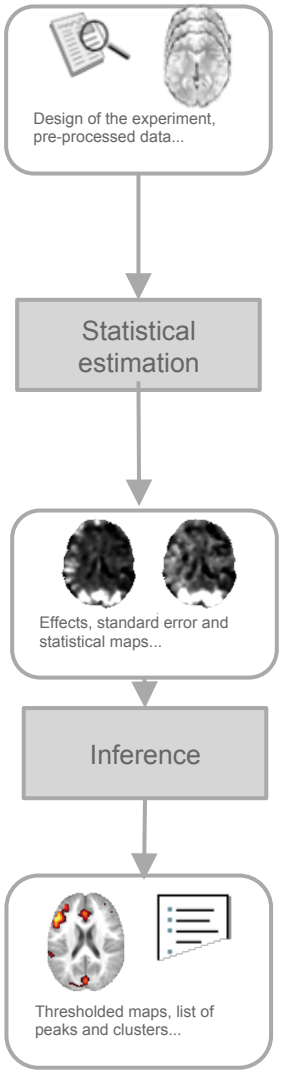
Semantic Web

- PROV-DM
  - Semantic web model for provenance
- NIDM
  - PROV-DM for neuroimaging

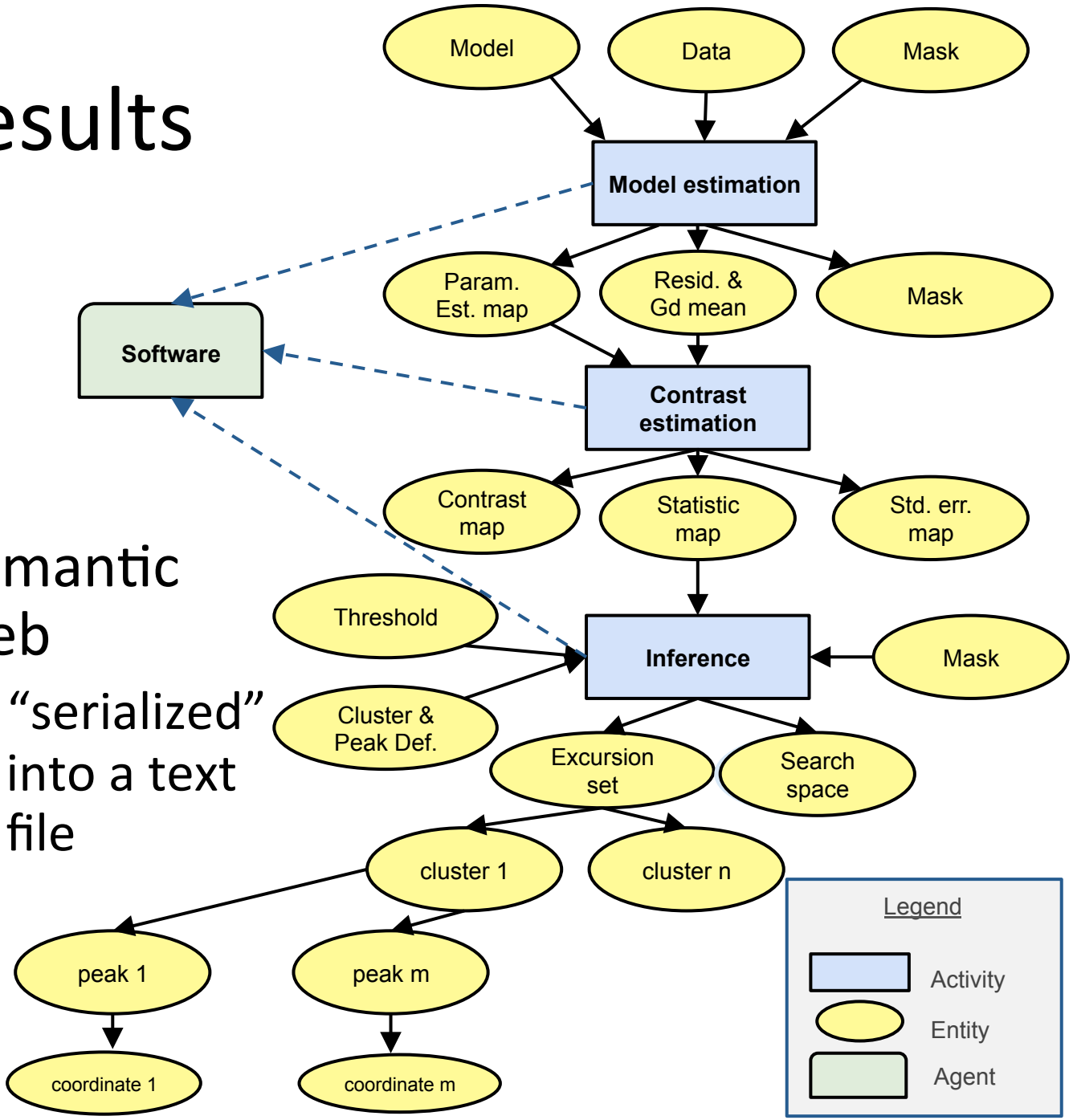




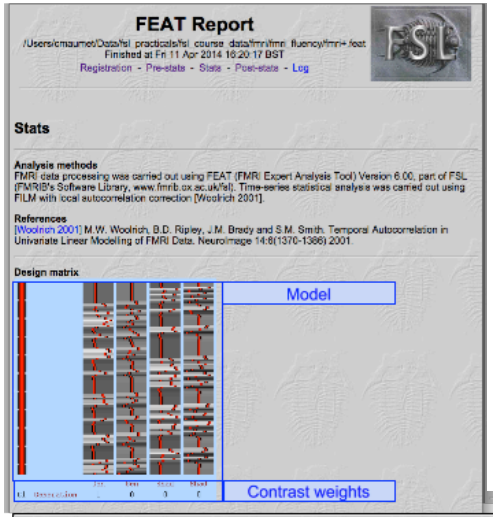
NIDM Results

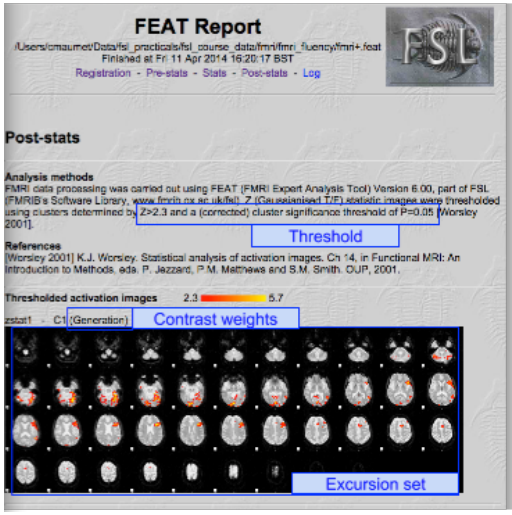


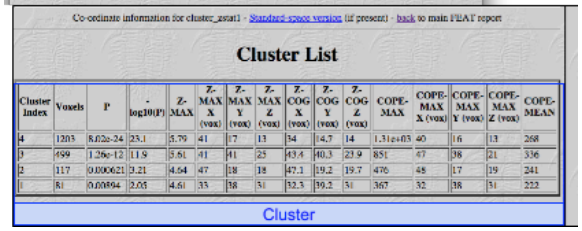
- Semantic web
  - "serialized" into a text file

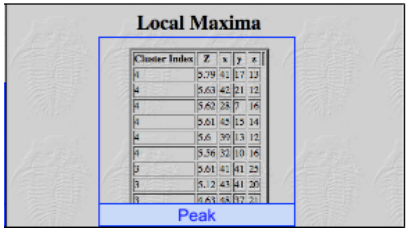


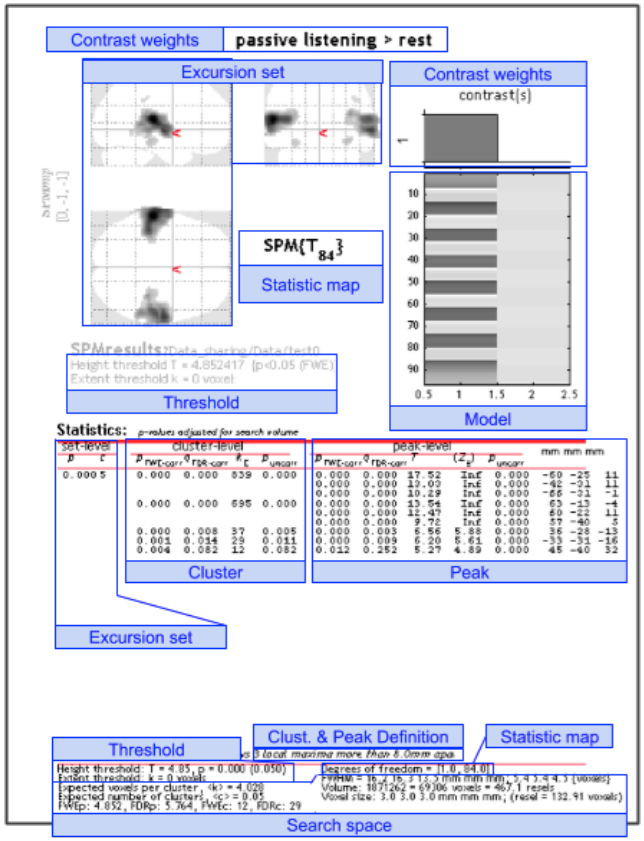
Common Format for all software

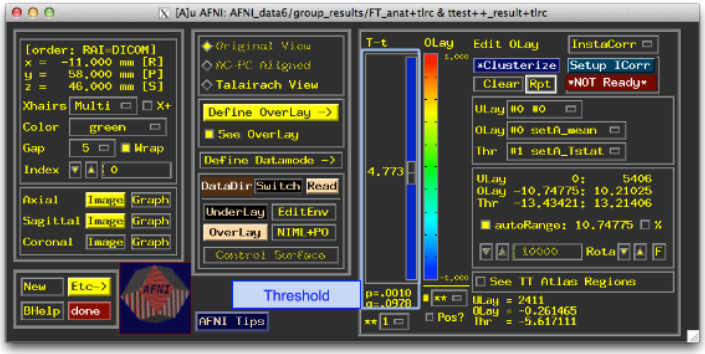


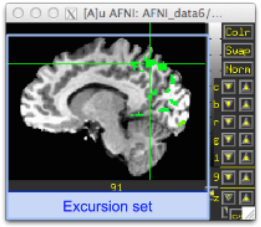


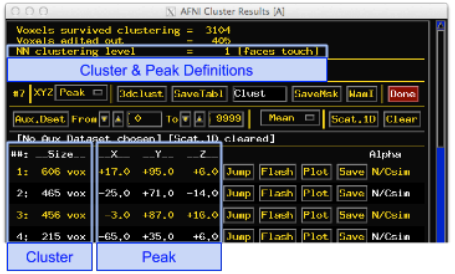












NIDM Exporters

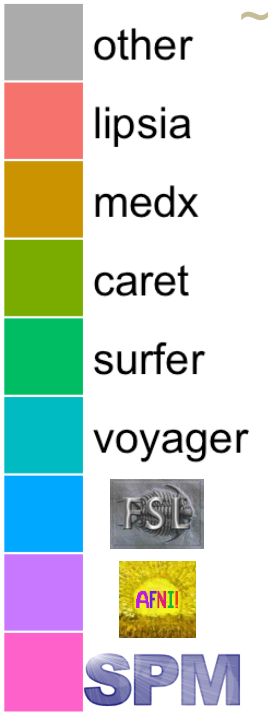
- SPM12 done!

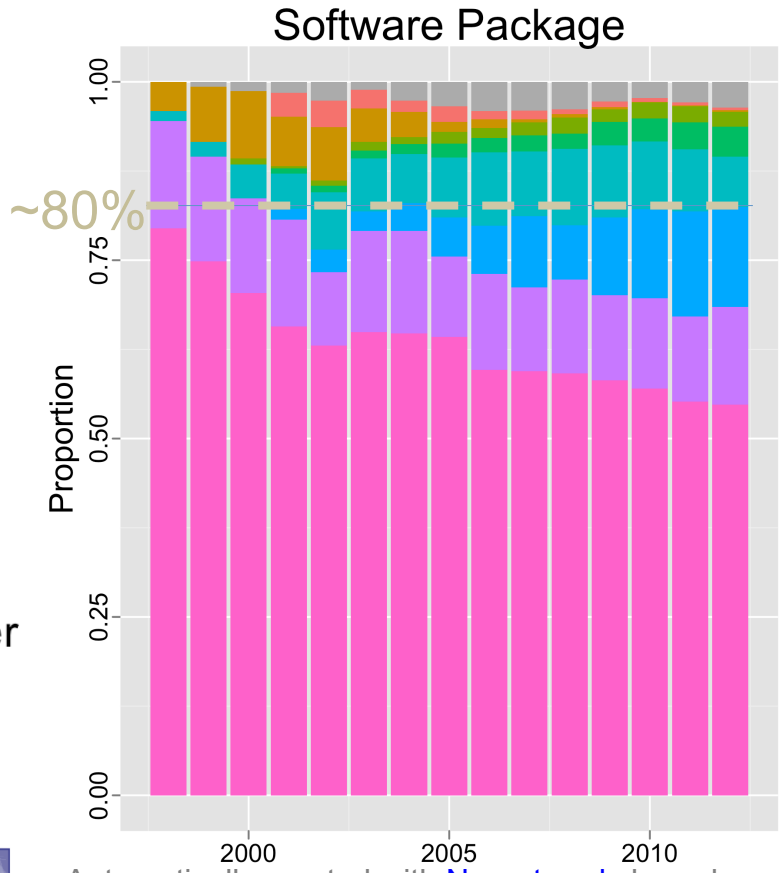
  SPM 12 batch system: SPM ➔ Stats ➔ Results report

  – SPM8 extension underway
- FSL – in beta testing
- AFNI

  – In planning stage
- For more see

  http://nidm.nidash.org



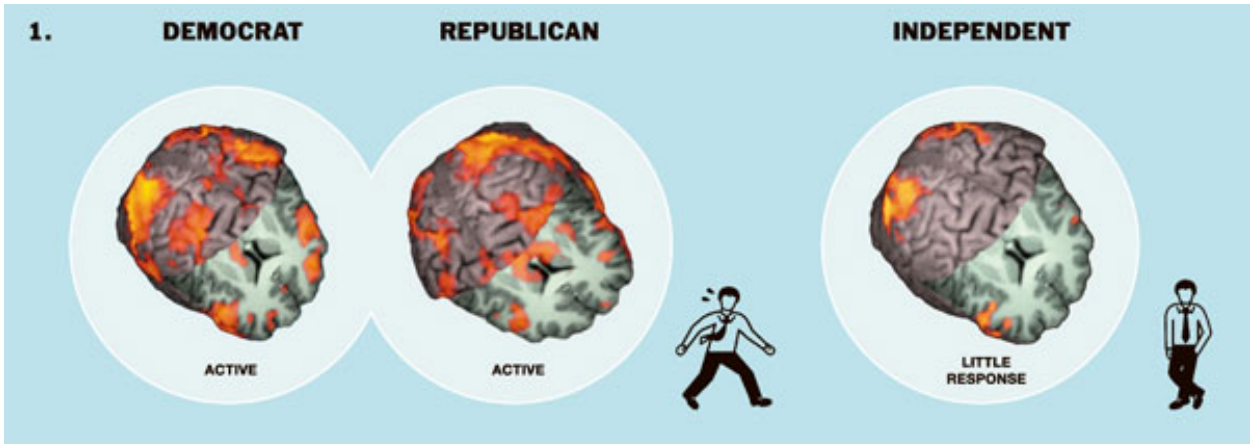


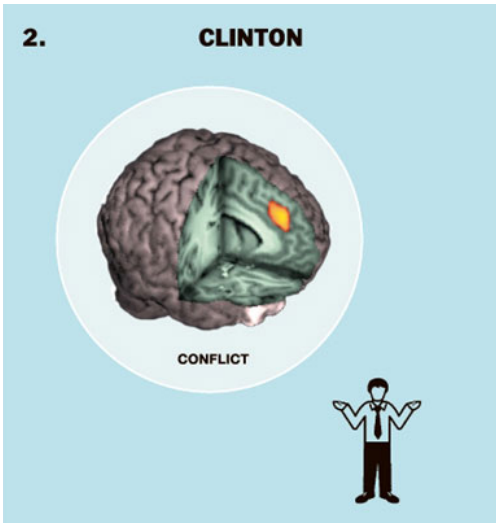
Automatically created with Neurotrends based on over 16 000 journal articles;

Neurosynth & Reverse Inference

Reverse Inference & Brain Imaging

- Politics study from 2007
  - Voters viewed images of Democratic candidates (N=20)
  - Subset that disliked Clinton:
    - "…exhibited significant activity in the anterior cingulate cortex, an emotional center"…, activated when one "feels compelled to act in two different ways but must choose one."





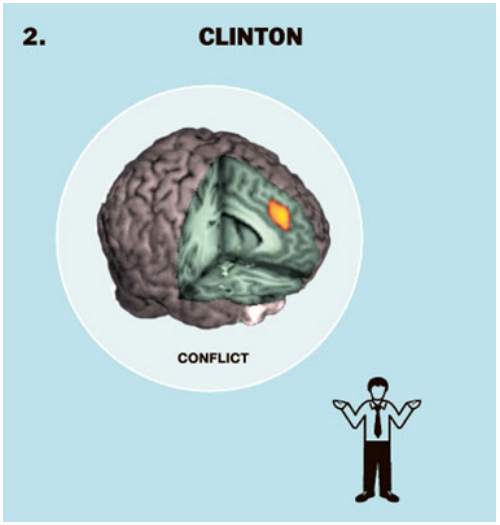
Iacoboni, et al., "This is your brain on politics". OP-ED, *The New York Times*, Nov. 11, 2007

Reverse Inference & Brain Imaging

- Logic
  - Emotion conflict resolution task
    ➔ Anterior Cingulate activation

    known from the literature
  - Hillary Clinton
    ➔ Anterior Cingulate activation

    observed in this experiment
  - Ergo
    ➔Hillary Clinton induces emotional conflict



Iacoboni, et al., *The New York Times*, Nov. 11, 2007

➔Faulty Reverse Inference

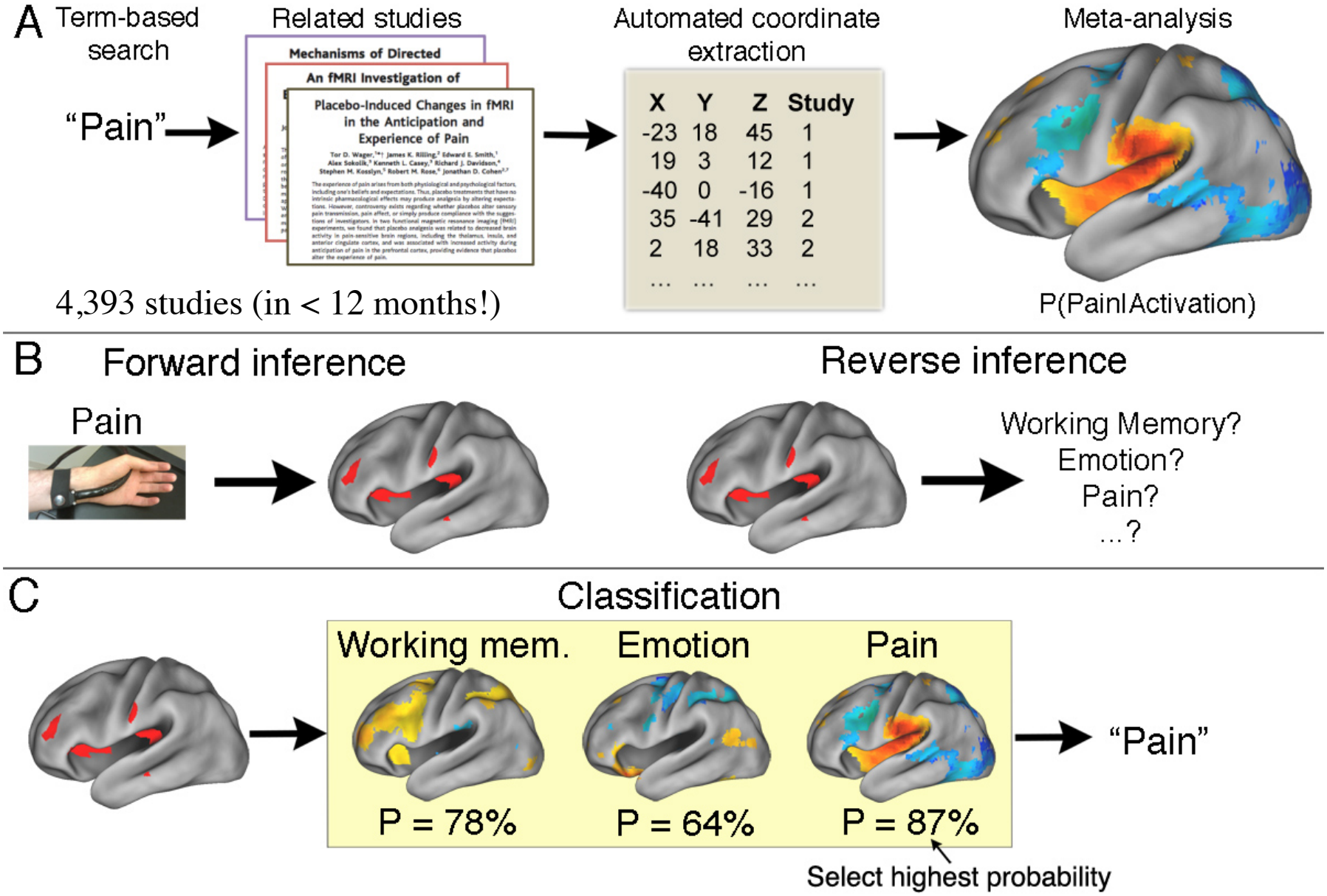
- High P(A.C. Act. | Emot. Conf.) *doesn't imply* high P(Emot. Conf. | A.C. Act.) *!!!*

Reverse Inference: Correctly!

- Bayes Rule
  - Cognitive Domain C, Activation A
  - $P(C=c|A) = P(A|C=c)\,P(C=c) \;/\; \Sigma_{c^*} P(A|C=c^*)P(C=c^*)$

    summation over *all cognitive domains!*
- Can we find "P(Emot. Conflict | ACC Act.)"?
  - Need to run 100's of experiments!
  - Or, use meta analysis!
  - But best Neuroimaging Meta Analysis databases are still limited
    - BrainMap.org has 2,757 studies (started in 1988)
    - Pubmed finds 28,694 refs "fMRI" in title/abstract

Neurosynth



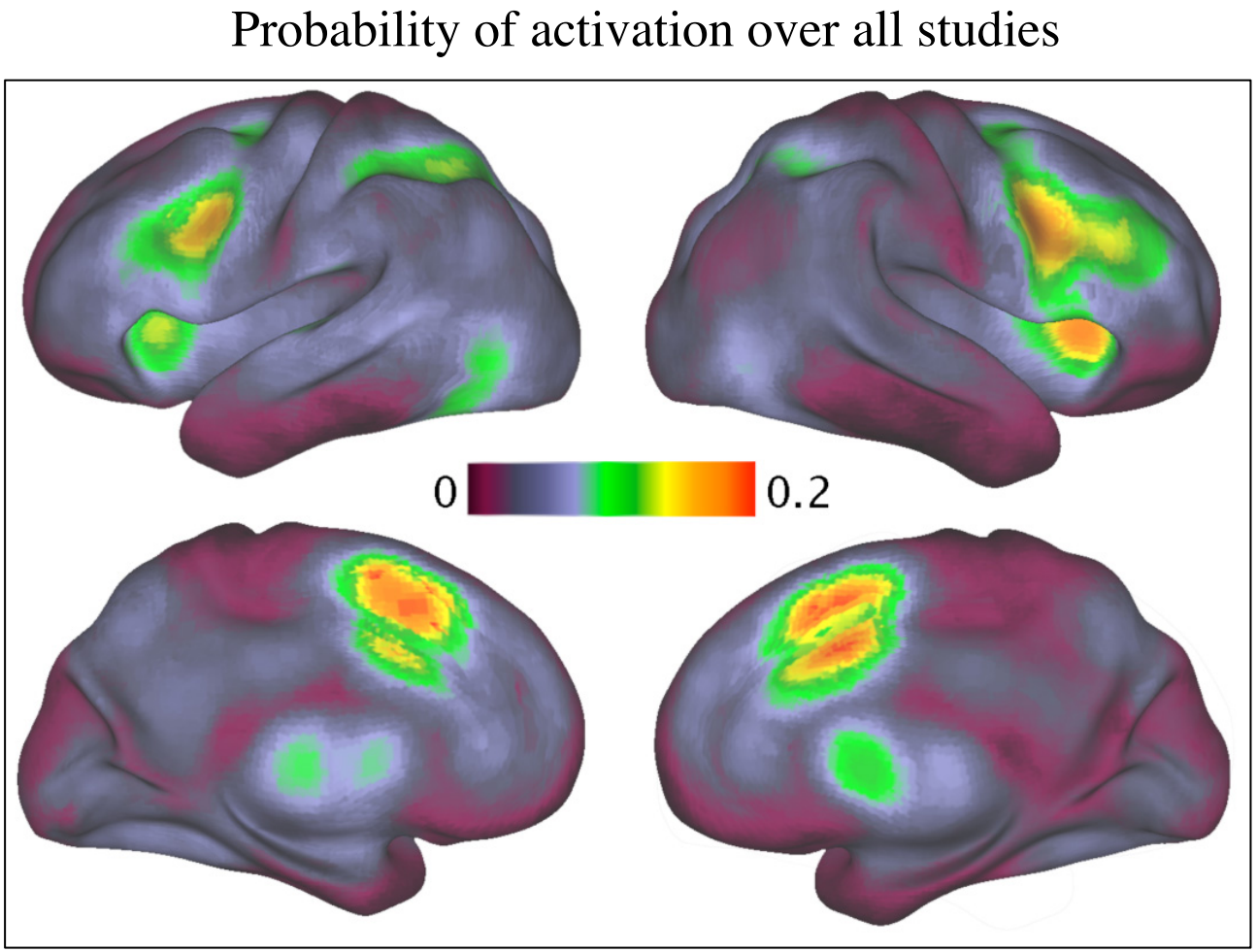
Yarkoni, Poldrack, Nichols, Essen, & Wager (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8), 665-670. www.neurosynth.org

Neurosynth Methods

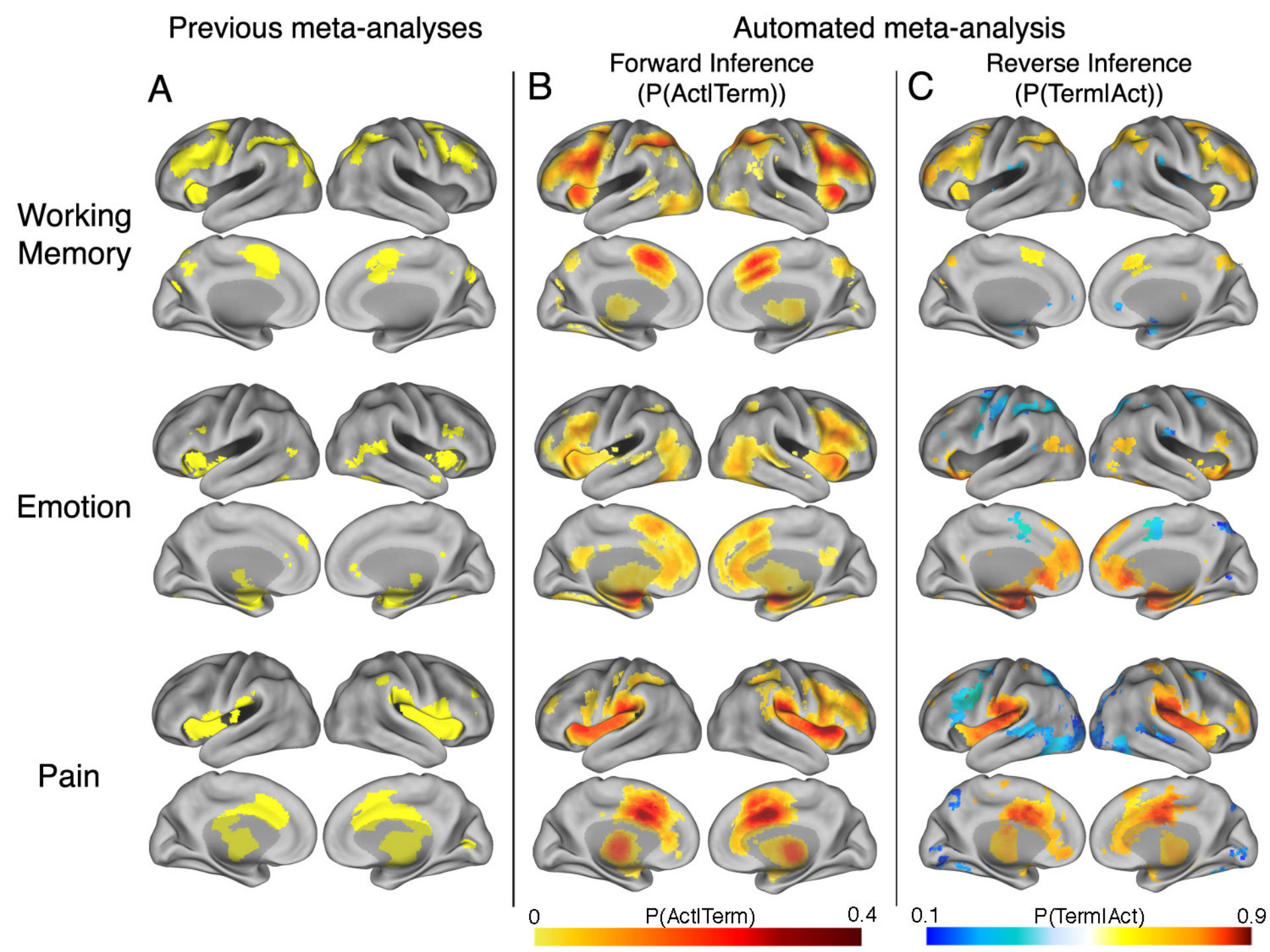
- 17 Neuroscience-focused journals used
  - *Biological Psychiatry, Brain, Brain and Cognition, Brain and Language, Brain Research, Cerebral Cortex, Cognitive Brain Research, Cortex, European Journal of Neuroscience, Human Brain Mapping, Journal of Neurophysiology, Journal of Neuroscience, NeuroImage, NeuroLetters, Neuron, Neuropsychologia*, & *Pain*.
- Tagging
  - Each article 'tagged' with psychological terms
  - Scored as high frequency (>1/1000 words) or not
- Coordinate harvesting
  - Tables parsed for x,y,z coordinates
- Not exhaustive, but already massive
  - 4,400+ studies, 145,000+ foci

What about Anterior Cingulate?

- It's always there!



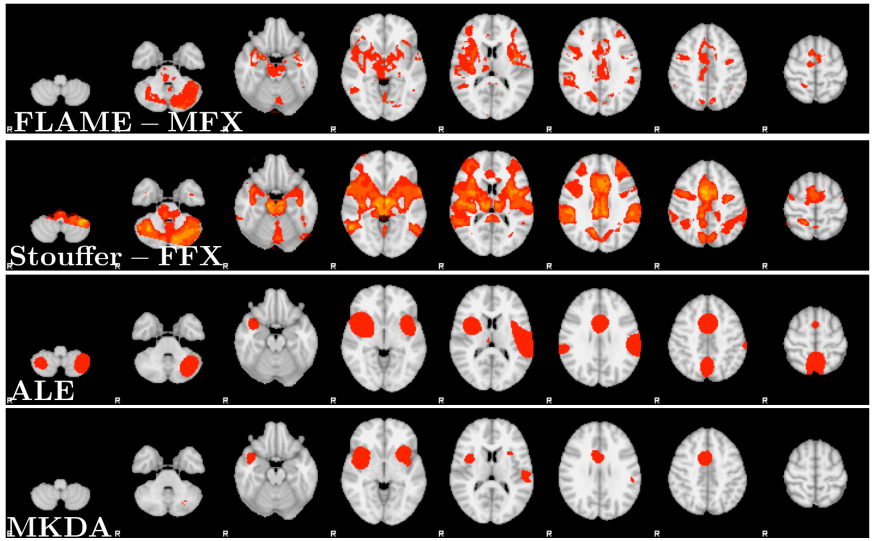
- Finally, can do real reverse inference…

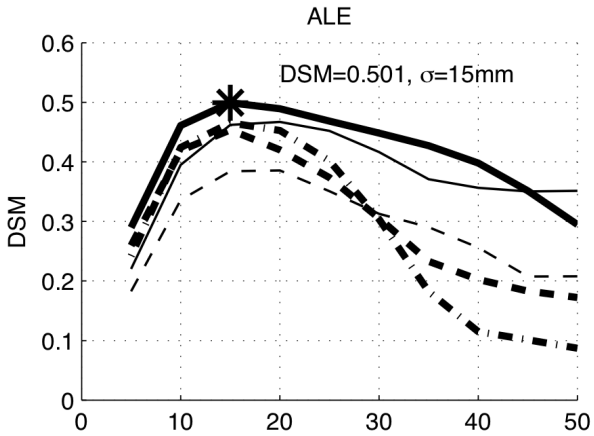

Previous meta-analyses
Automated meta-analysis
A
B
Forward Inference
(P(ActlTerm))
C
Reverse Inference
(P(TermlAct))
Working Memory
Emotion
Pain
0
P(ActlTerm)
0.4
0.1
P(TermlAct)
0.9

NeuroVault

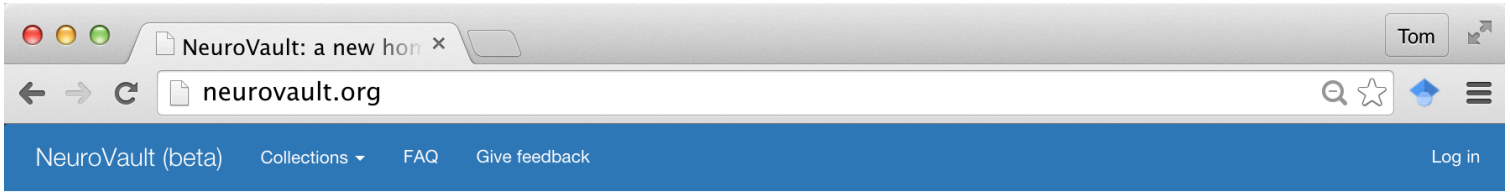
NeuroVault

- BrainMap | Neurosynth
  - Only coordinates
  - Huge loss of information
- Coordinate based vs. Intensity based Meta Analysis
  - Substantial information loss





Salimi-Khorshidi, G., Smith, S. M., Keltner, J. R., Wager, T. D., & Nichols, T. E. (2009). Meta-analysis of neuroimaging data: A comparison of image-based and coordinate-based pooling of studies. NeuroImage, 45(3), 810–823.



NeuroVault

A public repository of unthresholded statistical maps of the human brain

What is it?

A place where researchers can publicly store and share unthresholded statistical maps produced by MRI and PET studies.

Why use it?

- Interactive visualization
- A permanent URL
- Publicly shareable
- Improves meta-analyses

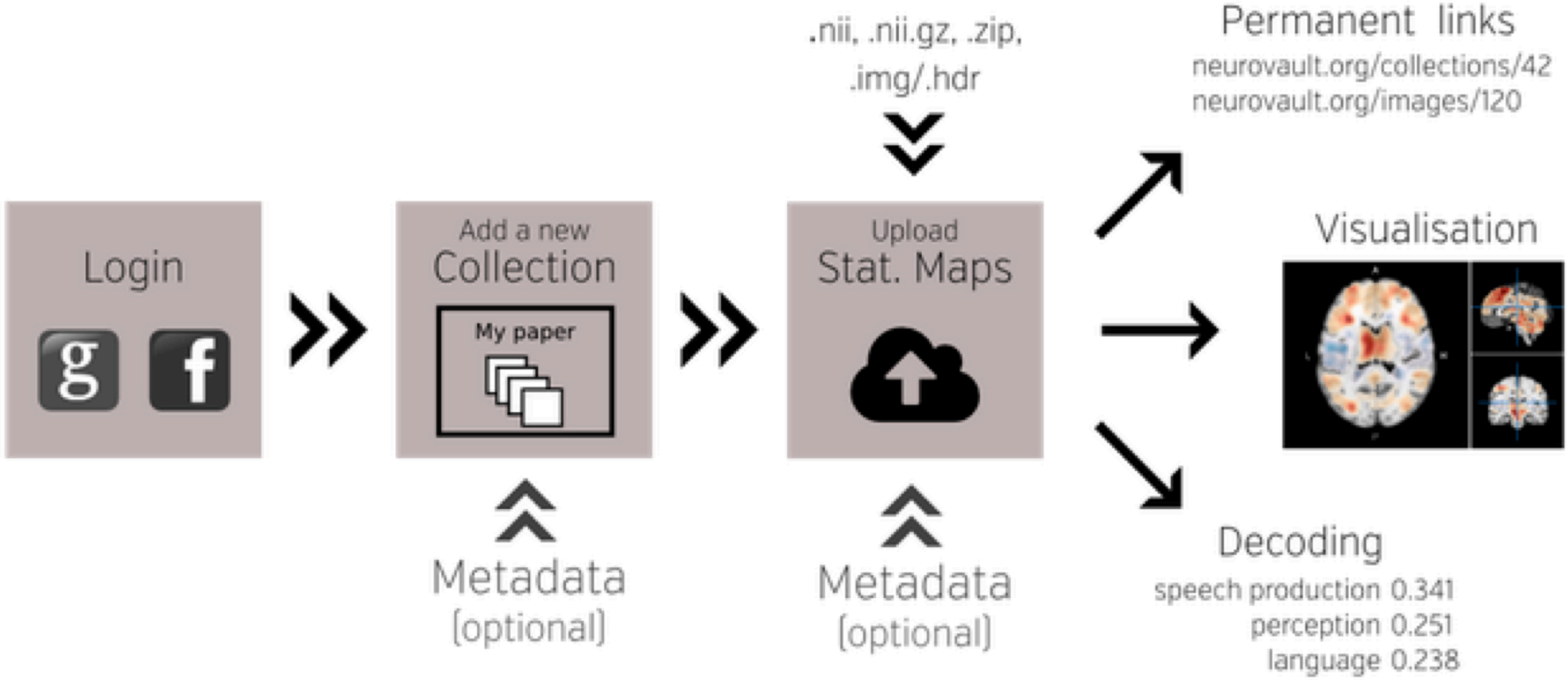
Supported by

Get started and upload an image!

Recently added collections of images from published papers

Name	Number of images
Physical Activity Is Linked to Greater Moment-To-Moment Variability in Spontaneous Brain Activity in Older Adults	100
Intrinsic visual-motor synchrony correlates with social deficits in autism	10
The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices	3
Altered Brain Activation during Emotional Face Processing in Relation to Both Diagnosis and Polygenic Risk of Bipolar Disorder	16
Patients with focal cerebellar lesions show reduced auditory cortex activation during silent reading	11

NeuroVault.org



Thank you!