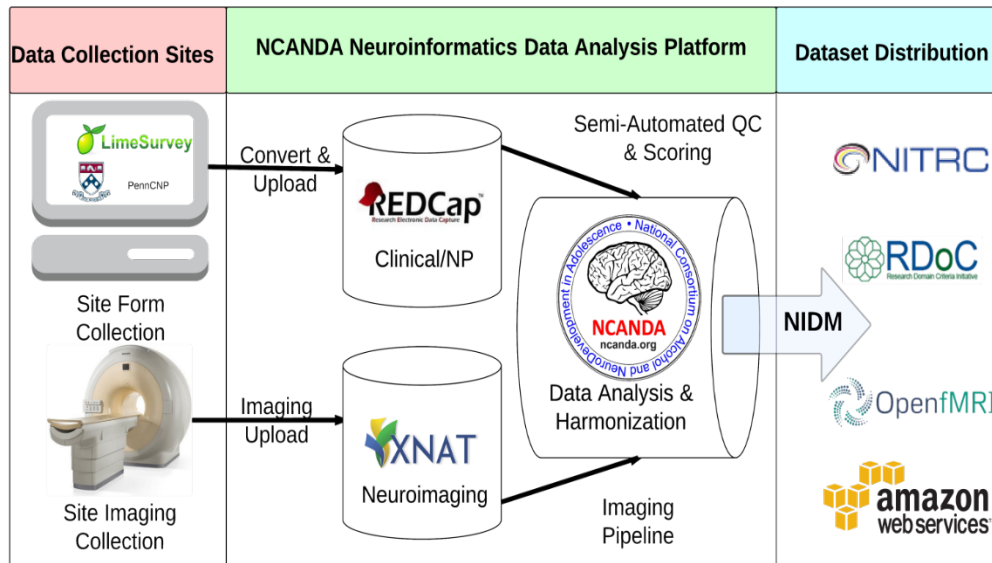


# Curating, Releasing, and Access NCANDA Data



**Kilian M. Pohl**  
Adolf Pfefferbaum

Ehsan Adeli  
Dongjin Kwon  
Michael Hasak  
Qingyu Zhao  
Sara Benito  
Weiwei Chu

SRI International

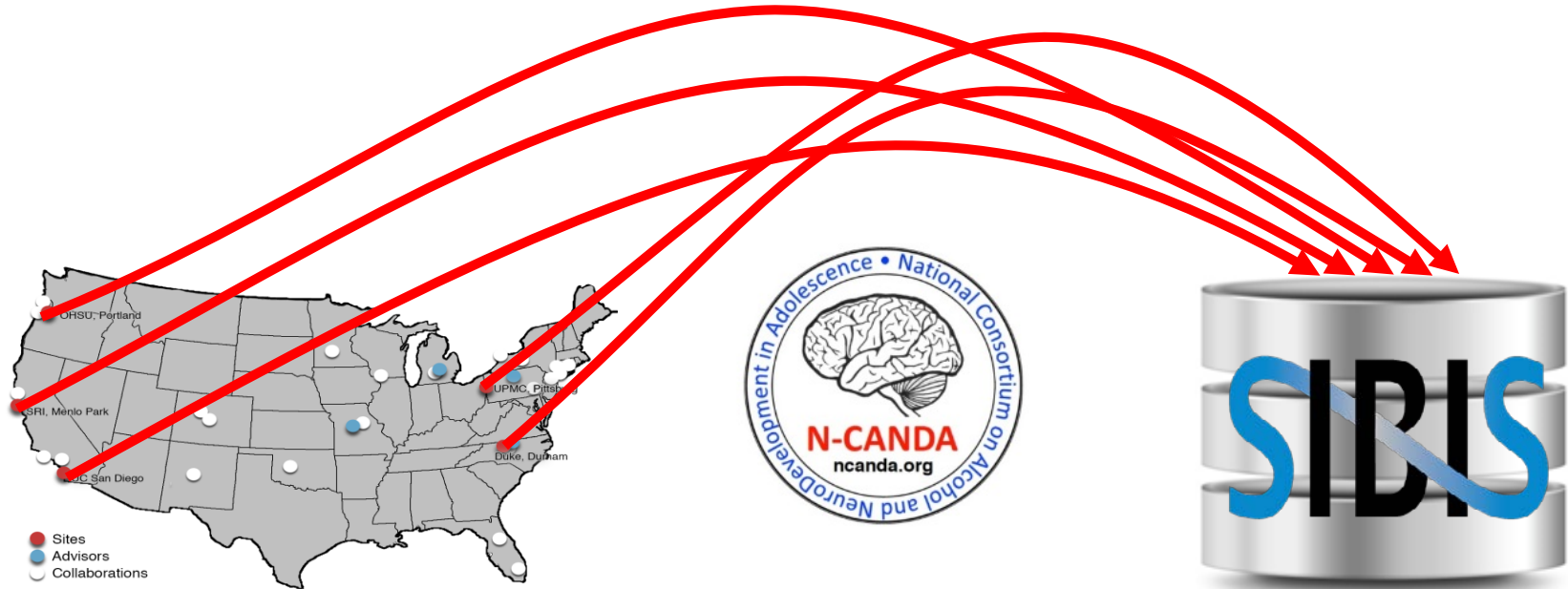
Edith V. Sullivan  
Stanford University

# Financial Interest Disclosure

Salary and Research Support  
of  
Kilian M Pohl, Ph.D.

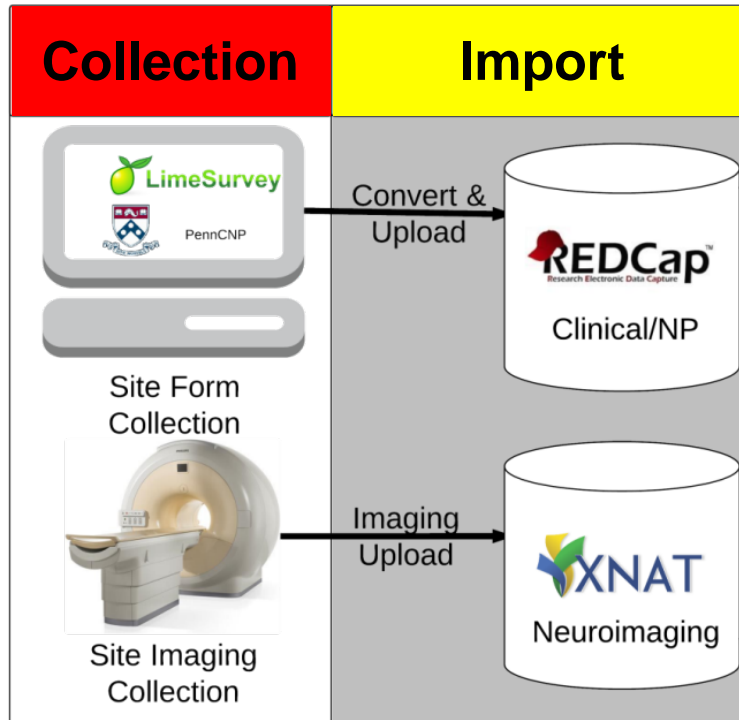


# Scalable Informatics for Biomedical Imaging Studies



- NCANDA collection sites upload data acquired in 831 study participants to the SIBIS portal
- SIBIS combines, curates, analyzes, and distributes the data using open source software.

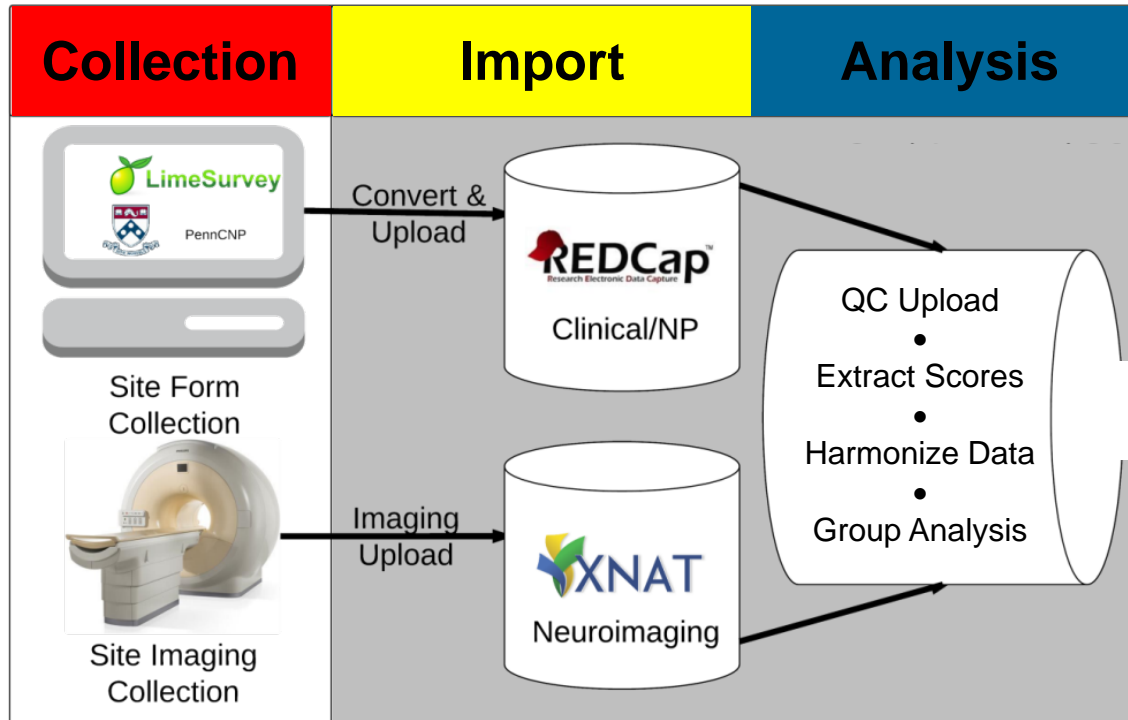
# Scalable Informatics for Biomedical Imaging Studies



## Sites collect:

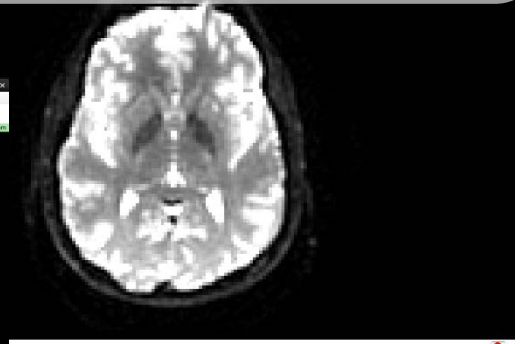
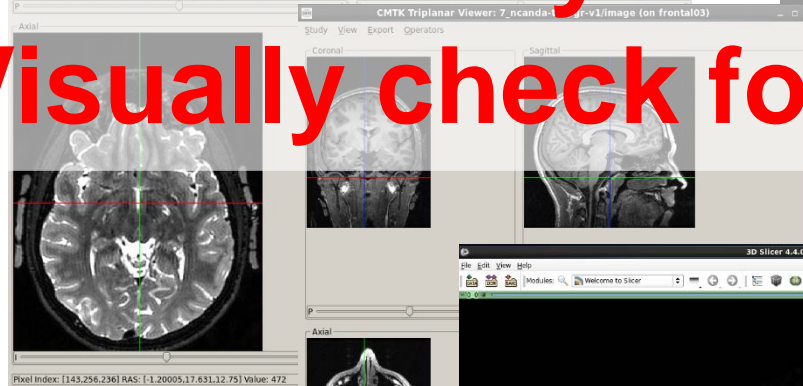
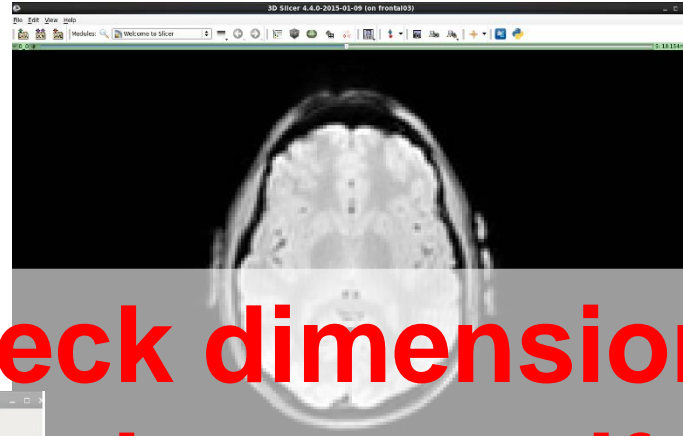
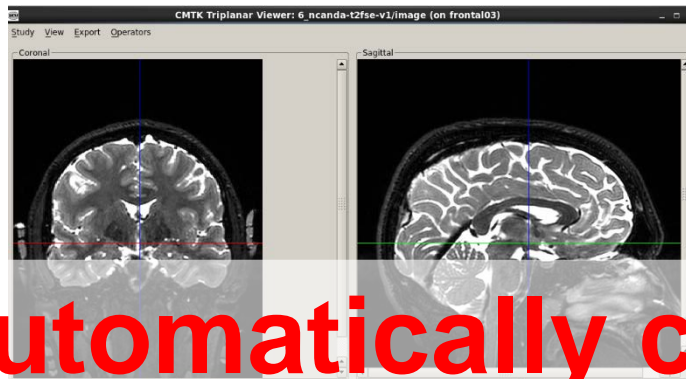
- Non Imaging Data
  - Demographic Information
  - Clinical Data
  - Neuropsychological Test Scores
- MRI
  - Structural
  - Diffusion
  - Functional

# Data Analysis



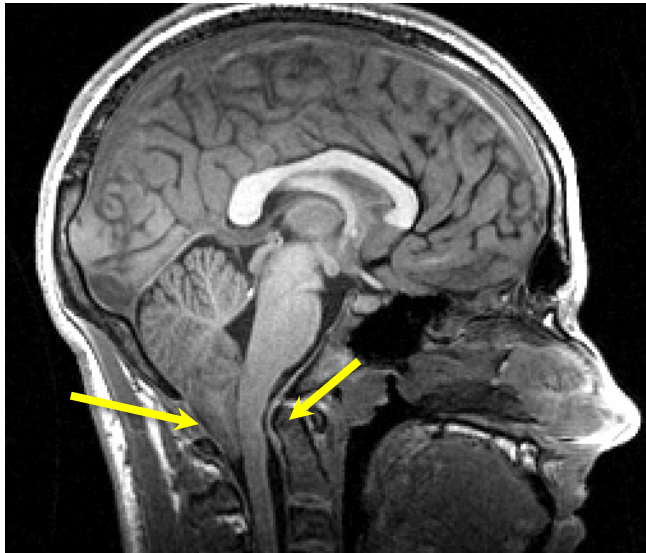
# Semi-Automatic Quality Control

**Automatically check dimension, ....  
Visually check for image artifacts**



# Neuroradiologist Reading

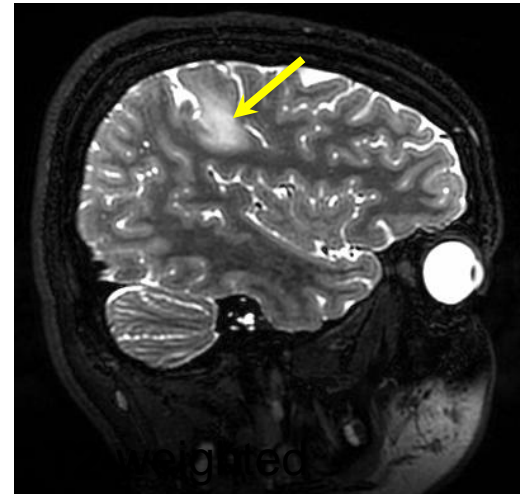
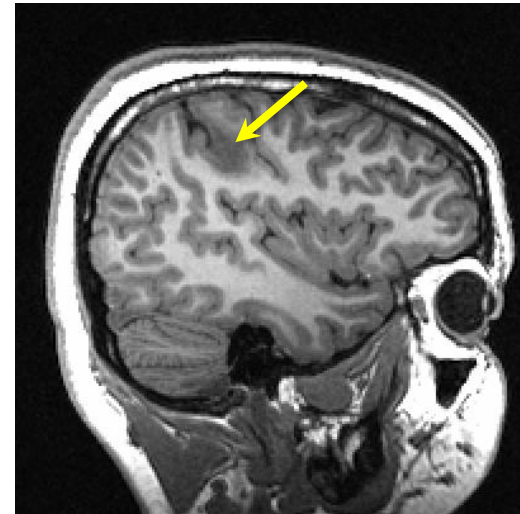
Chiari 1 malformation



bilateral tonsillar herniation  
with medullary distortion

- 95/833 adolescents = 11.4%
- 2 excluded from the study

right parietal cortical mass



# Issues Reported To Github

sibis-platform / ncanda-operations PRIVATE

Watch 3 Star 0 Fork 0

Code Issues 70 Pull requests 0 Wiki Pulse Graphs Settings

Filters is:issue is:open Labels Milestones New issue

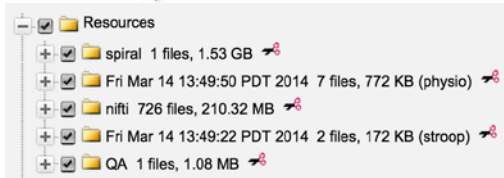
<input type="checkbox"/>	70 Open	✓ 1,524 Closed	Author	Labels	Milestones	Assignee	Sort
<input type="checkbox"/>	<b>D-00389-F-3-20160527, ERROR: Incorrect number of frames.</b> <a href="#">check_new_sessions</a>						0
	#1594 opened 16 hours ago by sri-ncanda						
<input type="checkbox"/>	<b>D-00142-M-6-20160526, ERROR: Incorrect number of frames.</b> <a href="#">check_new_sessions</a>						0
	#1593 opened 16 hours ago by sri-ncanda						
<input type="checkbox"/>	<b>C-70135-F-8, Missing MRI for Subject with visit date.</b> <a href="#">import_mr_sessions</a>						0
	#1592 opened 20 hours ago by sri-ncanda						
<input type="checkbox"/>	<b>NCANDA_E05448, QA metric fails to meet threshold.</b> <a href="#">phantom_qa</a>						0
	#1591 opened a day ago by sri-ncanda						
<input type="checkbox"/>	<b>C-70177-M-9, Missing MRI for Subject with visit date.</b> <a href="#">import_mr_sessions</a>						0
	#1590 opened 2 days ago by sri-ncanda						
<input type="checkbox"/>	<b>C-70135-F-8, No MR session for Subject C-70135-F-8 between 2016-01-30 and 2016-05-29</b>						0
	<a href="#">import_mr_sessions</a> #1589 opened 2 days ago by sri-ncanda						

Issues are resolved with collection sites





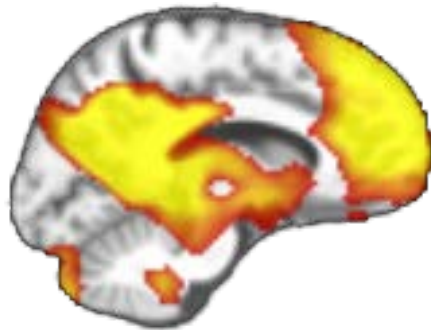
# Process Data



Import to  
Pipeline

Light Weight Data Pipeline  
<http://www.nitrc.org/projects/lwdp>

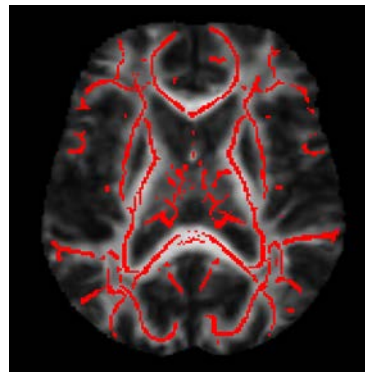
Resting State



Nipype-based  
Preprocessing

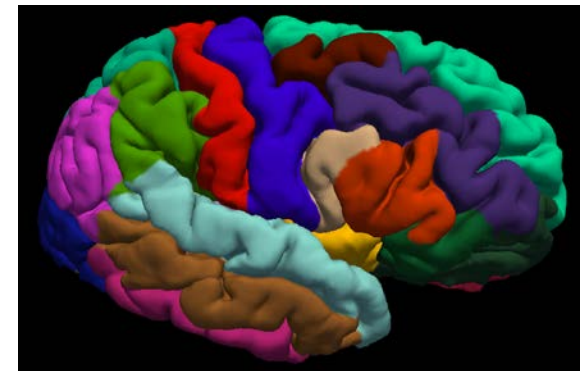
Kilian M. Pohl

Diffusion



FSL Tract-  
Based Spatial  
Statistics

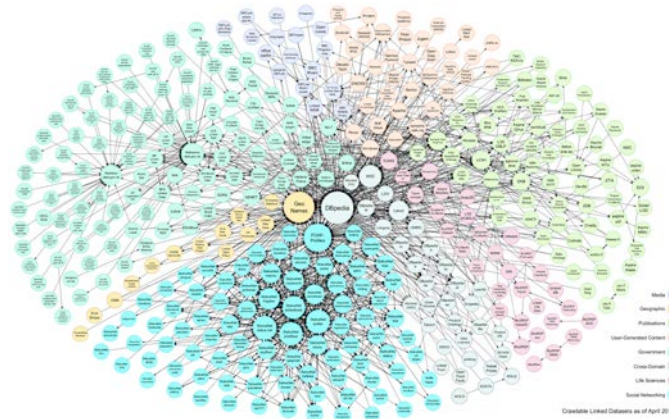
Anatomical



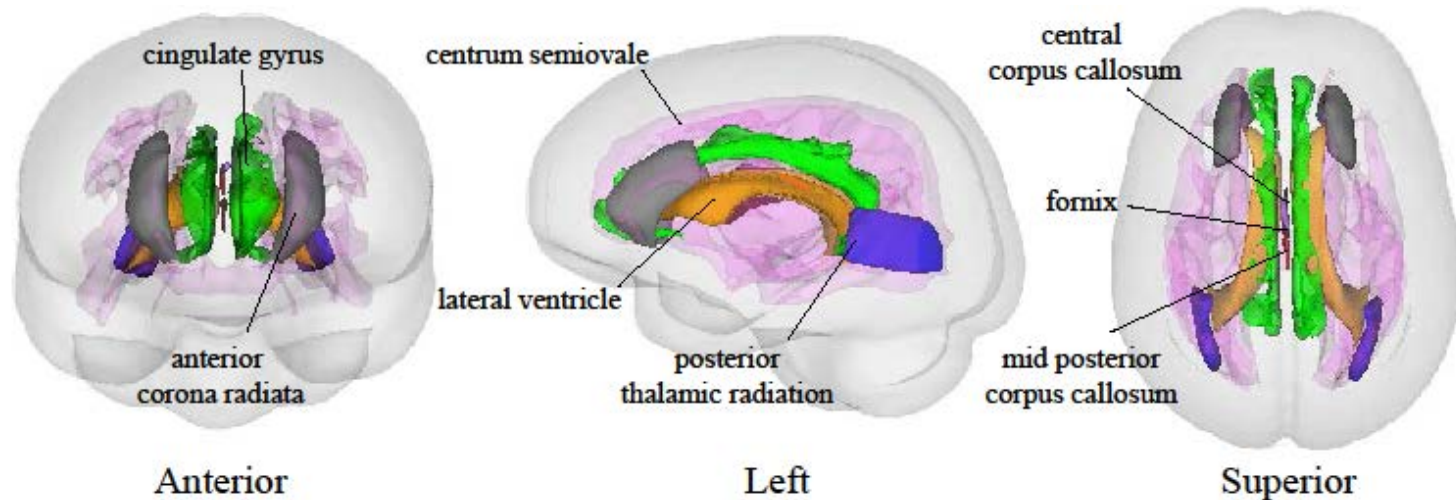
Volume Scores

# Create Data Releases

- Contains text files with de-identified demographic and neuropsychological test scores, raw and derived imaging data, and composite scores
- Data dictionary describing variables
- Software used to generate the data

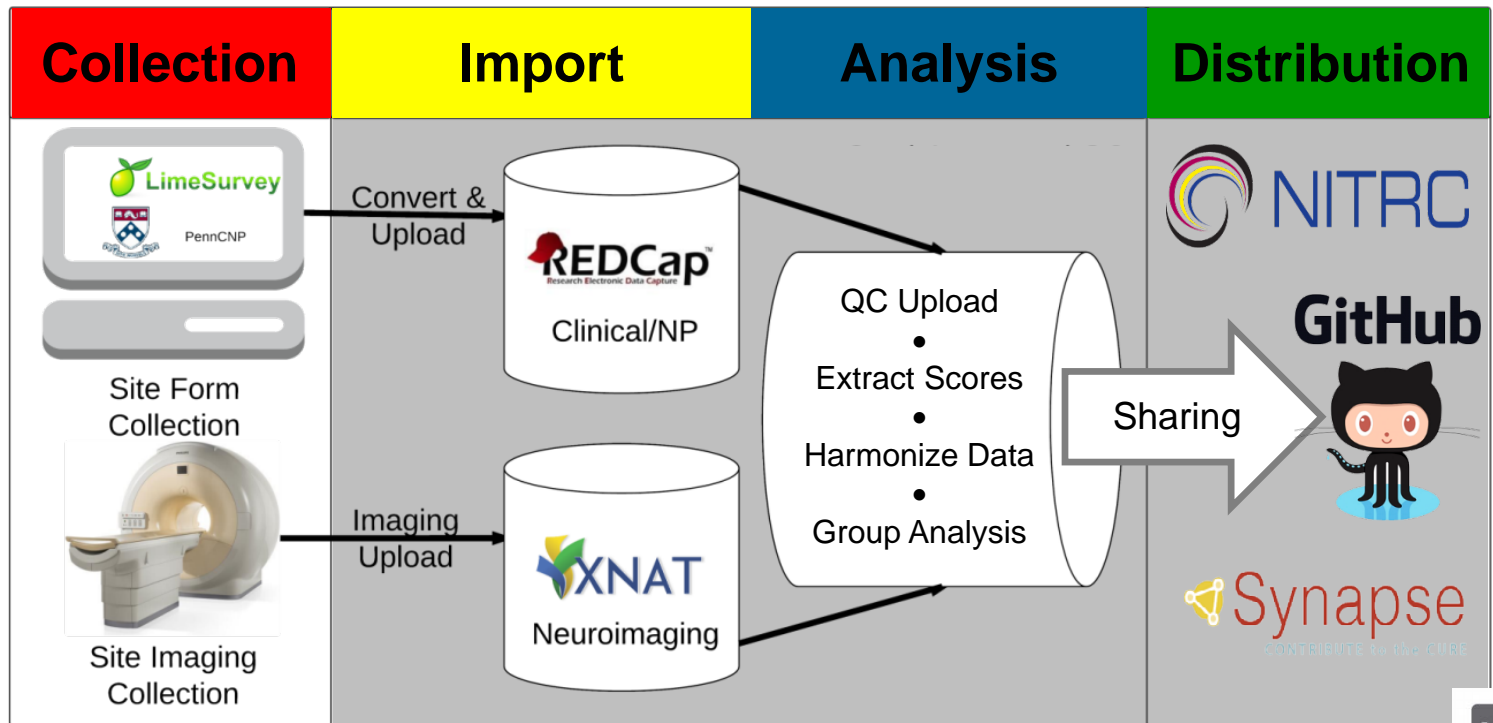


# Group Analysis



Create machine learning technology to identify brain regions impacted by regular alcohol consumption during adolescence

# Share Data and Software



Upload results to the data repository Synapse

Store software via

- NITRC: <https://www.nitrc.org/projects/ncanda-datacore>
- Github: <https://github.com/sibis-platform>



# Gaining Access to Data

Apply to NIAAA:

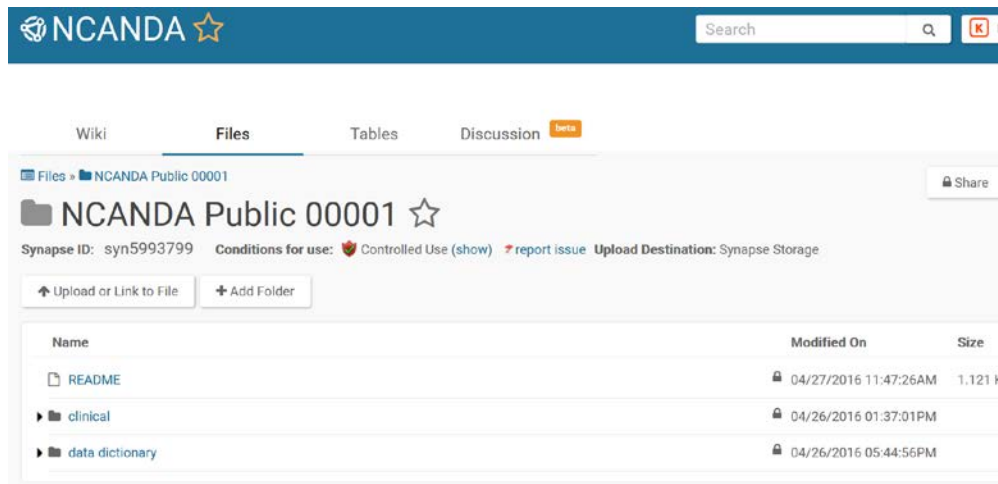
- A **cover letter** on the letterhead of the sponsoring institution at which the research study will be conducted.
- **Curriculum vitae** of the principal investigator and all co-investigators
- A 1-2 page **description of the proposed research**
- A completed **Data Distribution Agreement** with signatures of the principal investigator and an authorized representative of the sponsoring institution.

Information available at <https://www.niaaa.nih.gov/research/major-initiatives/national-consortium-alcohol-and-neurodevelopment-adolescence>

# Gaining Access to Data

After approval from NIAAA

- SRI will request from PI their Synapse user name
- PI will be able to access data via <https://www.synapse.org/ncanda>



The screenshot displays the Synapse web interface for the NCANDA Public 00001 dataset. The page features a blue header with the NCANDA logo and a search bar. Below the header, there are navigation tabs for Wiki, Files, Tables, and Discussion. The main content area shows the dataset name 'NCANDA Public 00001' with a star icon, a Synapse ID of 'syn5993799', and a 'Controlled Use' status. A table lists the dataset's contents, including a README file and two folders: 'clinical' and 'data dictionary'.

Name	Modified On	Size
README	04/27/2016 11:47:26AM	1.121 KE
clinical	04/26/2016 01:37:01PM	
data dictionary	04/26/2016 05:44:56PM	

Synapse allows users to query and download data releases, which are linked to publications.

# Data Sample

Wiki ? Files ? Tables ? Discussion ? Do

Files » NCANDA Public 00001 » clinical » clinical.csv

clinical.csv ☆ Download other download options

Download file as: clinical.csv | 333.099 KB | Synapse Storage md5 ?

Synapse ID: syn5993808 Conditions for use: Controlled Use (show) report issue

Preview ?

subject	arm	visit	aeq_gpc	aeq_csb	aeq_icma	aeq_se	aeq_cmi
NCANDA_S00033	standard	baseline	3.0	3.0	3.0	5.0	15.0
NCANDA_S00034	standard	baseline	8.0	8.0	5.0	9.0	13.0
NCANDA_S00042	standard	baseline	7.0	5.0	7.0	9.0	15.0
NCANDA_S00043	standard	baseline	9.0	9.0	9.0	9.0	9.0
NCANDA_S00051	standard	baseline	5.0	5.0	3.0	5.0	14.0
NCANDA_S00052	standard	baseline	5.0	7.0	5.0	8.0	13.0

Discussion about clinical.csv

Score

## Data Dictionary

	A	B	C
1	Variable / Field Name	Form Name	Field Label
2	aeq_gpc	clinical	Global Positive Change
3	aeq_csb	clinical	Changes in Social Behavior
4	aeq_icma	clinical	Improved Cognitive and Motor Ability
5	aeq_se	clinical	Sexual Enhancement
6	aeq_cmi	clinical	Cognitive and Motor Impairment
7	aeq_ia	clinical	Increased Arousal
8	aeq_rtr	clinical	Relaxation and Tension Reduction
9	aeq_mean	clinical	Mean Alcohol Non-Drinking Expectancies
10	aeq_pos	clinical	Positive Alcohol Non-Drinking Expectancies
11	aeq_neg	clinical	Negative Alcohol Non-Drinking Expectancies
12	casq_score	clinical	Cleveland Adolescent Sleepiness Score
13	cesd_score	clinical	CES Symptomatology Score
14	ctq_ea	clinical	Emotional Abuse Scale Total Score
15	ctq_pa	clinical	Physical Abuse Scale Total Score
16	ctq_sa	clinical	Sexual Abuse Scale Total Score
17	ctq_en	clinical	Emotional Neglect Scale Total Score
18	ctq_pn	clinical	Physical Neglect Scale Total Score
19	ctq_minds	clinical	Minimization/Denial Scale Total Score
20	drhq_tmv	clinical	Total Number of Moving Violations/Accidents
21	drhq_dss	clinical	Driving Safety Score
22	drhq_dui	clinical	Sum for Driving Under the Influence Score
23	hss_pastweek	clinical	HSS Past Week Score
24	hss_pastmonth	clinical	HSS Past Month Score



# Thank You



Dongjin Kwon, PhD



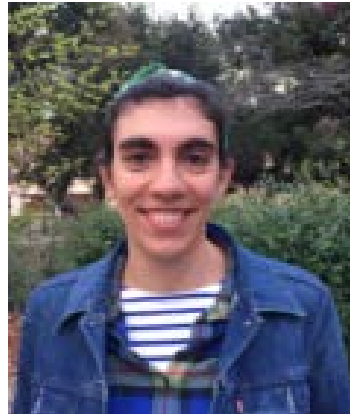
Ehsan Adeli, PhD



Qingyu Zhao, PhD



Michael Hasak



Sara Benito



Weiwei Chu