

Situating Discovery in the Data Sharing Environment

The Efforts and Goals of the Data Discovery Collaboration Project

Nicole Contaxis, MLIS
Lead, NYU Data Catalog
nicole.contaxis@nyulangone.org

The Problem

Data Discovery, ca. 2011

88% of datasets from
NIH-funded
publications were
'invisible' - between
200,000 and 235,000
datasets



RESEARCH ARTICLE

Sizing the Problem of Improving Discovery and Access to NIH-Funded Data: A Preliminary Study

Kevin B. Read^{1☯*}, Jerry R. Sheehan^{2☯}, Michael F. Huerta^{2☯}, Lou S. Knecht^{2☯}, James G. Mork^{2☯}, Betsy L. Humphreys^{2☯}, NIH Big Data Annotator Group^{3¶}

1 Medical Library, NYU Langone Medical Center, New York, New York, United States of America, **2** National Library of Medicine, National Institutes of Health, Bethesda, Maryland, United States of America, **3** National Institutes of Health, Bethesda, Maryland, United States of America

☯ These authors contributed equally to this work.

¶ Membership of the NIH Big Data Annotator Group is listed in the Acknowledgments.

* kevin.read@nyumc.org



Data Discovery, Now

- More repositories and other tools available to help researchers share data
- More requirements and policies that encourage data sharing
- Yet...

Data Discovery, Now

- More repositories and other tools available to help researchers share data
 - More requirements and policies that encourage data sharing
 - Yet...
- Complex environment with thousands of repositories (2,433 listed in re3data)
 - Difficult to decide where to deposit one's data or look for data for re-use
 - Some data (e.g., EHR data) may be useful but will not be included in repositories
-

The Technology

The NYU Data Catalog: An Overview

- Provides a standardized metadata schema to describe data

NYU HEALTH SCIENCES LIBRARY

NYU Data Catalog

Neurological Emergencies Outcomes at NYU

NYU Dataset

Alternate Title(s): NEON
UID: 10330
Author(s): Ariane Lewis*, Aaron Lord
* Corresponding Author

Description

This dataset was collected as part of a combined retrospective and prospective cross-sectional study to establish risk factors for infection after intracerebral hemorrhage and subarachnoid hemorrhage and to determine the impact of those infections on long-term outcomes. Data was harvested from Tisch Hospital records from January 2013 to December 2014 retrospectively and from January 2015 to the present prospectively, and the study aims to recruit an additional 1,000 patients by 2027.

Patients are included in the study if they are over 18 years of age and have a new diagnosis of intracerebral hemorrhage or subarachnoid hemorrhage requiring admission to or consultation by acute neurology faculty members at NYU Langone Medical Center, and for prospective patients, if the patient or next of kin consent to participate in follow-up phone interviews at 3 months and 12 months.

Data that will be collected from both retrospectively and prospectively enrolled patients include:

- Admission data (hospital admission information, history of present illness)
- Admission vital signs (BMI, weight, height, temperature, heart rate, respiratory rate, blood pressure)
- Admission labs (serum chemistries, blood count, coagulation)
- Baseline data (demographics, medications, past medical history, social history, family history)
- Admission examination (Hunt/Hess grade, Glasgow Coma Scale (GCS), NIH Stroke Scale (NIHSS), premorbid Modified Rankin Scale (MRS)
- Admission CT scan and angiogram results
- Hospital procedures, surgical treatments, medical treatments

Access via Data Request Form
Form to request access

Access Restrictions

Application Required
Author approval required

Access Instructions

Please contact Dr. Ariane Lewis for information on how to apply for access to this dataset.

Data Type

[Administrative](#)
[Clinical Measures](#)
[Imaging](#)
[Interviews](#)

Study Type

Observational

Dataset Format(s)

SPSS, Stata, Microsoft Excel, CSV

Data Collection Instruments

Glasgow Outcome Scale
Modified Rankin Scale
Barthel Index
Neuro-QOL

The NYU Data Catalog: An Overview

- Provides a standardized metadata schema to describe data
- Makes research data discoverable regardless of where it is stored

NYU HEALTH SCIENCES LIBRARY

NYU Data Catalog

Neurological Emergencies Outcomes at NYU

NYU Dataset

Alternate Title(s): NEON
UID: 10330
Author(s): Ariane Lewis*, Aaron Lord
* Corresponding Author

Description

This dataset was collected as part of a combined retrospective and prospective cross-sectional study to establish risk factors for infection after intracerebral hemorrhage and subarachnoid hemorrhage and to determine the impact of those infections on long-term outcomes. Data was harvested from Tisch Hospital records from January 2013 to December 2014 retrospectively and from January 2015 to the present prospectively, and the study aims to recruit an additional 1,000 patients by 2027.

Patients are included in the study if they are over 18 years of age and have a new diagnosis of intracerebral hemorrhage or subarachnoid hemorrhage requiring admission to or consultation by acute neurology faculty members at NYU Langone Medical Center, and for prospective patients, if the patient or next of kin consent to participate in follow-up phone interviews at 3 months and 12 months.

Data that will be collected from both retrospectively and prospectively enrolled patients include:

- Admission data (hospital admission information, history of present illness)
- Admission vital signs (BMI, weight, height, temperature, heart rate, respiratory rate, blood pressure)
- Admission labs (serum chemistries, blood count, coagulation)
- Baseline data (demographics, medications, past medical history, social history, family history)
- Admission examination (Hunt/Hess grade, Glasgow Coma Scale (GCS), NIH Stroke Scale (NIHSS), premorbid Modified Rankin Scale (MRS)
- Admission CT scan and angiogram results
- Hospital procedures, surgical treatments, medical treatments

Access via Data Request Form
Form to request access

Access Restrictions

Application Required
Author approval required

Access Instructions

Please contact Dr. Ariane Lewis for information on how to apply for access to this dataset.

Data Type

[Administrative](#)
[Clinical Measures](#)
[Imaging](#)
[Interviews](#)

Study Type

Observational

Dataset Format(s)

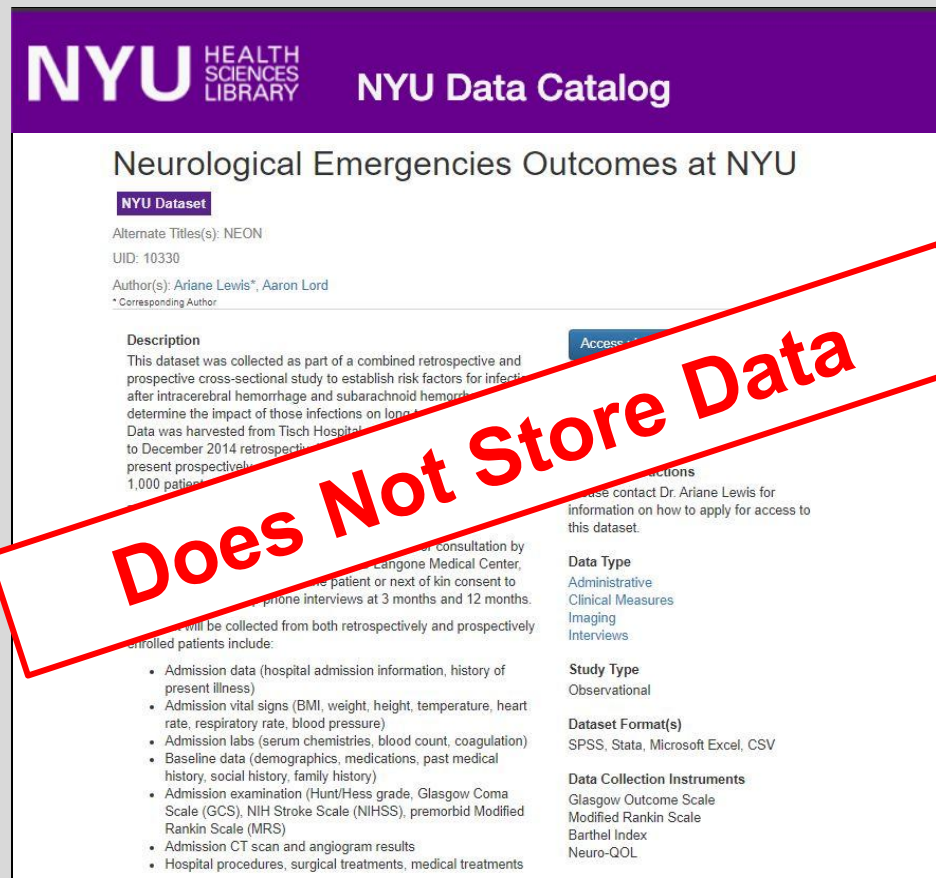
SPSS, Stata, Microsoft Excel, CSV

Data Collection Instruments

Glasgow Outcome Scale
Modified Rankin Scale
Barthel Index
Neuro-QOL

The NYU Data Catalog: An Overview

- Provides a standardized metadata schema to describe data
- Makes research data discoverable regardless of where it is stored



NYU HEALTH SCIENCES LIBRARY **NYU Data Catalog**

Neurological Emergencies Outcomes at NYU

NYU Dataset

Alternate Title(s): NEON
UID: 10330
Author(s): Ariane Lewis*, Aaron Lord
* Corresponding Author

Description
This dataset was collected as part of a combined retrospective and prospective cross-sectional study to establish risk factors for infection after intracerebral hemorrhage and subarachnoid hemorrhage. The study determined the impact of those infections on long-term outcomes. Data was harvested from Tisch Hospital from January 2010 to December 2014 retrospectively and prospectively. The study included 1,000 patients.

Access

Does Not Store Data

Data Type
Administrative
Clinical Measures
Imaging
Interviews

Study Type
Observational

Dataset Format(s)
SPSS, Stata, Microsoft Excel, CSV

Data Collection Instruments
Glasgow Outcome Scale
Modified Rankin Scale
Barthel Index
Neuro-QOL

- Admission data (hospital admission information, history of present illness)
- Admission vital signs (BMI, weight, height, temperature, heart rate, respiratory rate, blood pressure)
- Admission labs (serum chemistries, blood count, coagulation)
- Baseline data (demographics, medications, past medical history, social history, family history)
- Admission examination (Hunt/Hess grade, Glasgow Coma Scale (GCS), NIH Stroke Scale (NIHSS), premorbid Modified Rankin Scale (MRS)
- Admission CT scan and angiogram results
- Hospital procedures, surgical treatments, medical treatments

The NYU Data Catalog: An Overview

- Provides a standardized metadata schema to describe data
- Makes research data discoverable regardless of where it is stored
- Open source
 - Code on [GitHub](#)
 - Documentation on [OSF](#)

NYU HEALTH SCIENCES LIBRARY **NYU Data Catalog**

Neurological Emergencies Outcomes at NYU

NYU Dataset

Alternate Title(s): NEON
UID: 10330
Author(s): Ariane Lewis*, Aaron Lord
* Corresponding Author

Description
This dataset was collected as part of a combined retrospective and prospective cross-sectional study to establish risk factors for infection after intracerebral hemorrhage and subarachnoid hemorrhage. The study determined the impact of those infections on long-term outcomes. Data was harvested from Tisch Hospital from January 2010 to December 2014 retrospectively and from January 2015 to present prospectively. 1,000 patients were enrolled.

Access

Questions
Please contact Dr. Ariane Lewis for information on how to apply for access to this dataset.

Data Type
[Administrative](#)
[Clinical Measures](#)
[Imaging](#)
[Interviews](#)

Study Type
Observational

Dataset Format(s)
SPSS, Stata, Microsoft Excel, CSV

Data Collection Instruments
Glasgow Outcome Scale
Modified Rankin Scale
Barthel Index
Neuro-QOL

- Admission data (hospital admission information, history of present illness)
- Admission vital signs (BMI, weight, height, temperature, heart rate, respiratory rate, blood pressure)
- Admission labs (serum chemistries, blood count, coagulation)
- Baseline data (demographics, medications, past medical history, social history, family history)
- Admission examination (Hunt/Hess grade, Glasgow Coma Scale (GCS), NIH Stroke Scale (NIHSS), premorbid Modified Rankin Scale (MRS))
- Admission CT scan and angiogram results
- Hospital procedures, surgical treatments, medical treatments

An Extremely Abridged History, Part One

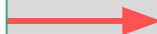
The NYU Data Catalog

- Increase visibility of research data for researchers at NYU
- Encourage collaboration

An Extremely Abridged History, Part One

The NYU Data Catalog

- Increase visibility of research data for researchers at NYU
- Encourage collaboration



The Data Catalog Collaboration Project

- Improve data discovery through institutional data catalogs
- Work together to develop, implement, and improve the Data Catalog software

Collaboration Members

- NYU Langone Health
- University of Pittsburgh
- University of Maryland at Baltimore
- Weill Cornell Medicine
- Wayne State University
- Zucker School of Medicine at Hofstra/Northwell
- Memorial Sloan Kettering Cancer Center
- The University of North Carolina at Chapel Hill
- Duke University
- University of Virginia

Storage Agnostic Data Discovery

Data Catalog
A project of the Health Sciences Libraries



LoginAbout the CatalogHome

Search here to find large public and licensed datasetsGo!

Murine vagus nerve recordings after cytokine injection

UID: 10366

Author(s): Theodoros P. Zanos, Harold A. Silverman, Todd Levy, Tea Tsaava, Emily Battinelli... [See more...](#)

* Corresponding Author

Description

These datasets were generated during a study to develop methods to isolate and decode specific neural signals to discriminate between two different cytokines. Nerve implants recorded from the vagus nerve of mice exposed to IL-1B and TNF were sorted into groups based on their shape and amplitude, and their respective firing rates were computed. These cytokine-mediated responses were subsequently decoded. Datasets contain all neurorecordings from the surface vagus nerve of mice. Matlab code for analysis and algorithms is included.

Subject of Study

Mice

Subject Domain

Electrophysiology

Keywords

Bioelectronic medicine
Cytokines
Inflammation
Neural decoding
Neural signals
Vagus nerve

Access via Feinstein Institutes for Medical Research repository

Access Instructions

Link to the local repository where the datasets and analysis code in Matlab reside is provided in the PMC article.

Associated Publications

Zanos TP, Silverman HA, Levy T, Tsaava T, Battinelli E, Lorraine PW, Ashe JM, Chavan SS, Tracey KJ, Bouton CE. Identification of cytokine-specific sensory neural signals by decoding murine vagus nerve activity. *Proc Natl Acad Sci U S A*. 2018 May 22;115(21):E4643-E4652. doi: 10.1073/pnas.1719003115. Epub 2018 May 7. PubMed PMID: 29735654; PubMed Central PMCID: PMC6003492.

Zucker School of Medicine at Hofstra/Northwell Academic Works Repository Zanos TP, Silverman HA, Levy T, Tsaava T, Battinelli E, Lorraine PW, Ashe JM, Chavan SS, Tracey KJ, Bouton CE. Identification of cytokine-specific sensory neural signals by decoding murine vagus nerve activity. . 2018 Jan 01; 115(21):Article 3087 [p]. Available from: <https://academicworks.medicine.hofstra.edu/articles/3087>. Free full text article

Data Type

Electrophysiological

Equipment Used

Bipolar sling platinum-iridium cuff electrode (CoTec)
Plexon data-acquisition system (OmniPlex; Plexon, Inc.)

Software Used

Matlab

Institutional Repository

Storage Agnostic Data Discovery

NYU HEALTH SCIENCES LIBRARY

NYU Data Catalog

HomeAbout the CatalogContact UsLogin

Search here to find large public and licensed datasets

Go!

Neurological Emergencies Outcomes at NYU

NYU Dataset

Alternate Title(s): NEON

UID: 10330

Author(s): Ariane Lewis*, Aaron Lord

* Corresponding Author

Description

This dataset was collected as part of a combined retrospective and prospective cross-sectional study to establish risk factors for infection after intracerebral hemorrhage and subarachnoid hemorrhage and to determine the impact of those infections on long-term outcomes. Data was harvested from Tisch Hospital records from January 2013 to December 2014 retrospectively and from January 2015 to the present prospectively, and the study aims to recruit an additional 1,000 patients by 2027.

Patients are included in the study if they are over 18 years of age and have a new diagnosis of intracerebral hemorrhage or subarachnoid hemorrhage requiring admission to or consultation by acute neurology faculty members at NYU Langone Medical Center, and for prospective patients, if the patient or next of kin

Access via Data Request Form

Form to request access

Access Restrictions

Application Required
Author approval required

Access Instructions

Please complete the Data Request Form to request access to this dataset. The information included in the form will be sent to the Corresponding Author who will evaluate your request for their data. The Corresponding Author may ask you to

Institutional Center or Department

Storage Agnostic Data Discovery

UNIVERSITY of MARYLAND
BALTIMORE

UMB Data Catalog
A project of the Health Sciences and Human Services Library

Find Datasets from
UMB Researchers

Go!

Include your dataset

Clinical Study Reports (CSR) Used in the Systematic Review of Neuraminidase Inhibitors for Preventing and Treating Influenza in Healthy Adults and Children with Confirmed or Suspected Exposure to Natural Influenza

[UMB Dataset](#)

Alternate Titles(s): CSRs used in Cochrane review of neuroaminidase inhibitors for influenza

UID: 10

Author(s): Peter Doshi*

* Corresponding Author

Description
Dataset consists of 107 full clinical study reports (CSRs) of published and unpublished randomized, placebo-controlled clinical trials of two neuraminidase inhibitors. A systematic review of the CSRs and other regulatory documents was conducted to determine the potential benefits and harms of oseltamivir (Tamiflu) and zanamivir (Relenza). Outcome measures included time to first alleviation of symptoms, influenza outcomes, complications, admissions to hospital, and adverse events in the intention to treat population. Completeness of relevant parts of the CSRs was determined via an extraction form based on the CONSORT statement checklist. Authors have provided the full set of clinical study reports for both medications provided to the Cochrane collaboration by Roche, GlaxoSmithKline, and the European Medicines Agency (EMA) for use in the systematic review of these neuraminidase inhibitors for treating/preventing influenza in healthy adults and children. A guest post on the "Dryad News and Views" site regarding the dataset of clinical study reports and the resulting Cochrane systematic review is available at: <https://blog.datadryad.org/2014/04/17/tamiflu-data/>

Subject of Study
Human

Subject Domain
Influenza, Human

Access via Dryad
Full set of clinical study reports and readme files


Access Restrictions
Free to All

Access Instructions
Available to download from the Dryad site

Associated Publications
Jefferson T, Jones M, Doshi P, Spencer EA, Onakpoya I., Heneghan CJ. (2014). Oseltamivir for influenza in adults and children: systematic review of clinical study reports and summary of regulatory comments. BMJ 348:g2545 (April 9). DOI: 10.1136/bmj.g2545
Heneghan CJ, Onakpoya I, Thompson M, Spencer EA., Jones M, Jefferson, T. (2014). Zanamivir for influenza in Adults and Children: Systematic Review of Clinical Study Reports and Summary of Regulatory Comments. BMJ 348:g2547 (April 9). DOI: 10.1136/bmj.g2547
Jefferson T, Jones MA, Doshi P, Del Mar CB, Thompson MJ, Spencer EA, Onakpoya IJ, Mahtani KR, Numan D, Howick J, Heneghan CJ. (2014).

External Repository

Storage Agnostic Data Discovery

 Get Data@Duke

Search here to find large public and licensed datasets

Go!

Brain Connectivity in AMD

Internal Dataset

UID: 4

Description

Includes 81 older adults with age-related macular degeneration (AMD) and 85 age-matched controls. All participants have comprehensive cognitive assessment, audiology, visual acuity, and demographic data. All eligible participants (about half of the cohort) underwent resting state brain MRI (fMRI, DTI, structural MRI). Mean age of cohort is 73.

Local Expert

[Heather Whitson](#)

Subject Domain

[Chronic Disease](#)

Population Age

Senior (65 years - 79 years)

Keywords

[Cognition](#)
[fMRI](#)
[Macular degeneration](#)
[MRI](#)

Access Restrictions

Available to Duke
IRB-approved protocol

Access Instructions

Please contact the Local Expert for further information: [Heather Whitson](#)

Associated Publications


Whitson HE, Chou Y, Potter GG, et al. Phonemic Fluency and Brain Connectivity in Age-Related Macular Degeneration: A Pilot Study. *Brain Connectivity*. 2015;5(2):126-135. doi:10.1089/brain.2014.0277.

Grant Support

R01 AG043438/NIH

Restricted Access Data

Storage Agnostic Data Discovery

 **WAYNE STATE**
UNIVERSITY

Data Catalog

Home About

Search here to find large public and licensed datasets

Go!

Adjunct Vitamin D Therapy as a Means to Reduce the Disparity in Subclinical Target Organ Cardiac Damage among Vulnerable Hypertensive Patients

WSU Dataset

Alternate Title(s): AddReach

UID: 6

Author(s): [Phillip Levy](#)

Description

The objective of this phase II/III randomized clinical trial is to evaluate the efficacy of vitamin D therapy versus placebo in vitamin D-deficient African-Americans with hypertension, including investigating the relationship between vitamin D and cardiac damage (as identified on cardiac magnetic resonance imaging) in a vitamin D-deficient hypertensive patients without prior history of heart disease. Data include cardiac MRI and echocardiography data, lab results (e.g., vitamin D, CBC), and patient characteristics (e.g., vital signs, demographics, health insurance, education level, household income, hypertension drug and vitamin D treatment adherence, dietary intake, sun exposure). ClinicalTrials.gov Identifier: [NCT01360476](#)

[Access via Author](#)

Access Restrictions

Application required

Access Instructions

Potential users can request access to the data by contacting the author.

Data Type

Clinical Measures
Survey

Dataset Format(s)

SAS, XLS

Dataset Size

2.5MB

Grant Support

1R01MD005849-01A1/National Institutes of Health

Timeframe

2011 - 2015

Subject Domain

[African Americans](#)
[Cardiovascular Health](#)
[Urban Issues](#)

Population Age

Adult (18 to 64 years)
Senior (65 to 79 years)

Keywords

Datasets Via Author Only

Storage Agnostic Data Discovery

NYU HEALTH SCIENCES LIBRARY

NYU Data Catalog

Search here to find large public and licensed datasets

Go!

Predicting Appointment Cancellation and No-show in Advance Using EHR: NYU Langone Health EHR NYU Dataset

UID: 10356

Description

This dataset was compiled as part of a project to better predict the probability of patient cancellation, no-show, and rescheduling of appointments in advance. For every appointment made between January 2016 and May 2018, this dataset includes: appointment unique ID, patient medical reference number, date and time of the appointment, the type of appointment (e.g., MRI, ultrasound, consult), appointment specialty (e.g., neurology, mammography), whether the appointment was attended by the patient in time, whether the patient did not attend the appointment, whether the patient rescheduled the appointment, whether the patient made the appointment themselves, the identification of the person who made the appointment and their relationship to the patient if the appointment was not made by the patient, the time and date that the appointment was made, the appointment doctor, the appointment copay with insurance, the appointment copay without insurance if applicable, the patient financial balance at the time of the appointment being made if available or applicable, the patient financial balance at the time of the appointment if applicable or available, the modality by which the appointment was made (e.g., by phone, in person), the unique ID of receptionist who made the appointment, whether the appointment was rescheduled, the date and time the patient asked for reschedule and the date and time for the new appointment for each time the appointment is rescheduled, the date and time the patient asked for a cancellation and the reason for the cancellation if the appointment was canceled, appointment reminder read status on MyChart, text message reminder sent on MyChart, the date and time of all MyChart logins, and patient demographics and zip code.

Timeframe

2016 - 2018

Geographic Coverage

New York (State) - New York City

Local Expert for NYU

[Narges Razavian](#)

Access via DataCore

Data Request Application
Accession #: 661

Access Restrictions

Application Required
NYU Langone Health Employees Only

Access Instructions

To request this dataset, please use the [DataCore form](#) and include Reference 661. Contact [DataCore](#) with any questions about pulling the data. Questions about using this dataset for research purposes can be directed to [Narges Razavian](#).

Data Type

[Clinical Measures](#)

Study Type

Observational

Dataset Format(s)

CSV

Electronic Health Record Data

Storage Agnostic Data Discovery

NYU HEALTH SCIENCES LIBRARY

NYU Data Catalog

Home

Search here to find large public and licensed datasets

Go!

[Edit this dataset](#)

Demographic, lesion pattern, and lesion location data for multiple sclerosis patients

NYU Dataset

UID: 10329

Author(s): Sanjeev Chawla, Ilya Kister*, Tim Sinnecker, Jens Querfel, Jean-Christophe Brisset... [See more...](#)

* Corresponding Author

Description

This dataset was compiled as part of an exploratory study to determine the longitudinal changes in multiple sclerosis (MS) lesions over time on ultra-high field MR imaging. Nine patients with MS underwent high-resolution 3D-susceptibility weighted imaging (SWI) and 2D-gradient-echo-T2*-weighted imaging on 7T MRI at baseline and after approximately 2.4 years for follow-up imaging. The data includes demographic image (patient gender, patient age, year of the onset of symptoms, the duration of the disease, the disease sub-type, ambulatory status, race, and years between baseline and follow-up imaging) as well as information on the pattern type and location of lesions at baseline and follow-up.

Geographic Coverage

New York (State) - New York City

Subject Domain

[Chronic Disease](#)
[Health Status](#)

Keywords

[Diagnostic imaging](#)
[MRI](#)
[Multiple Sclerosis](#)

[Access via PMC](#)

Supporting Information
Accession #: PMCID: 6136714

Access Restrictions

Free to All

Access Instructions

Data available for download through PubMed Central.

Associated Publications

Chawla S, Kister I, Sinnecker T, Wuerfel J, Brisset JC, Paul F, Ge Y. Longitudinal study of multiple sclerosis lesions using ultra-high field (7T) multiparametric MR imaging. *PLoS One*. 2018 Sep 13; 13(9):e0202918.

Data Type

[Clinical Measures](#)
[Physiological](#)

Study Type


[Observational](#)


Dataset Format(s)

[Microsoft Excel](#)

Datasets in Supplemental Files of Published Articles

Storage Agnostic Data Discovery

 University of Pittsburgh

 **Pitt Data Catalog**
A project by the Health Sciences Library System

Find datasets from University of Pittsburgh Researchers

Software: Neurophysiological analytics for all! Free open-source software tools for documenting, analyzing, visualizing, and sharing using electronic notebooks

UID: Z7

Author(s) Rosenberg, David M.; Horn, Charles C.**

* Corresponding Author
* University of Pittsburgh Author

Description

This Github repository contains raw data and code to explore open-source neurophysiology data analysis tools within an included Jupyter notebook. Software dependencies are listed within each data supplement folder, and can be downloaded with Docker via the associated Github repository described in the related data catalog record: [Software: Docker image with JupyterLab, Python 3, Python 2, and R](#). The supplementary data were collected from electrophysiological recordings of the musk shrew vagus, a model system to investigate gut-brain communication.

Keywords

[Electronic Lab Notebooks](#)
[Electrophysiology](#)
[Neurophysiology](#)
[Software](#)

Access via Github
Data files, full list of dependencies, and author-supplied instructions and information.

Access Restrictions
Free to all

Access Instructions
Download data files or access Binder link through Github. More information and instructions are available through the associated publication.

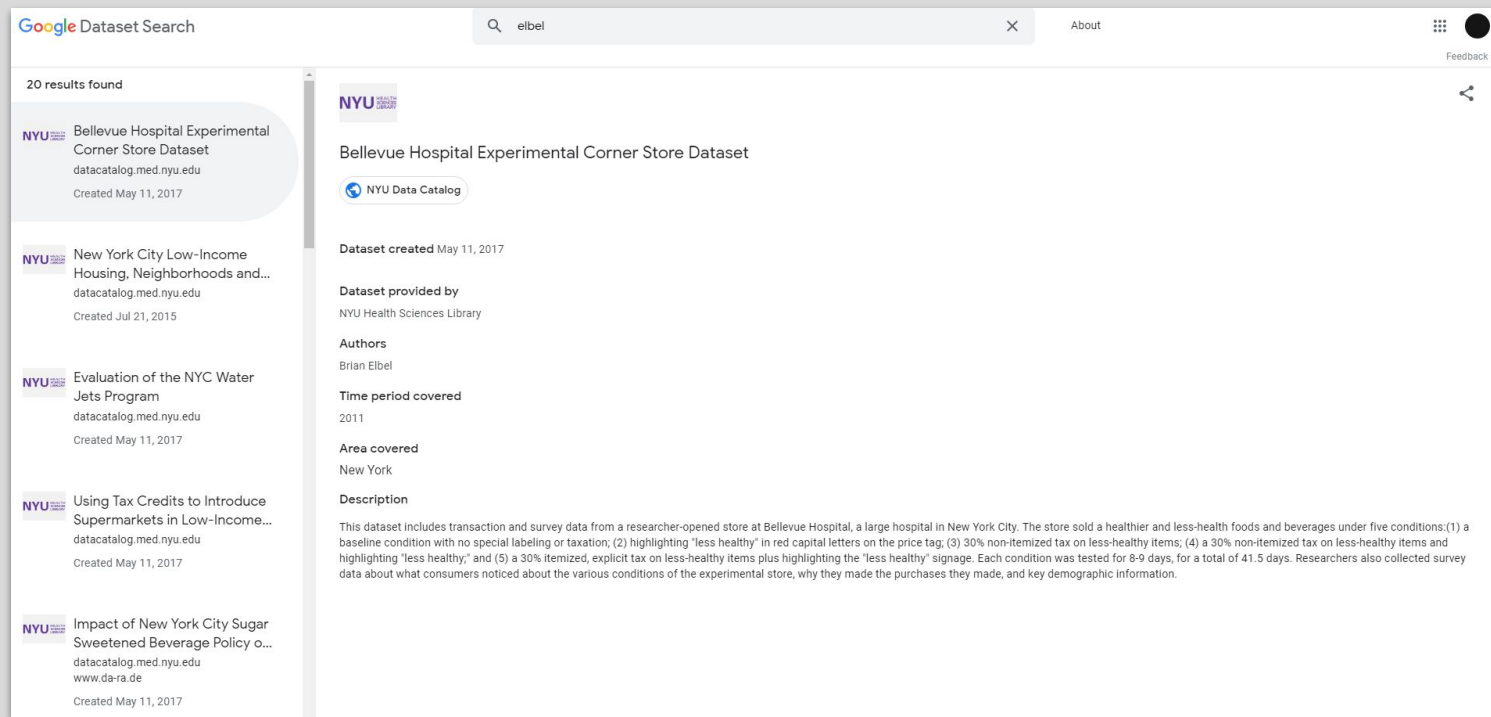
Associated Publications
Rosenberg DM, Horn CC. Neurophysiological analytics for all! Free open-source software tools for documenting, analyzing, visualizing, and sharing using electronic notebooks. J Neurophysiol. 2016 Aug 1;116(2):252-62.

Data Type
Electrophysiological

Software Used
ec
| R package for analysis of factorial experiments.
ggplot2
| R package for data visualization.
Jupyter Notebook
| App to create, edit, and run documents containing computer code and rich text.
Matplotlib
| Python 2D plotting library.
Neo
| Python module for loading and saving multiple electrophysiology data formats.
NumPy

Datasets with Associated Software

Interoperability with Schema.org



The screenshot shows the Google Dataset Search interface. At the top, the search bar contains 'elbel'. Below the search bar, it indicates '20 results found'. On the left, a list of results is shown, each with the NYU logo, a title, a URL, and a creation date. The first result is 'Bellevue Hospital Experimental Corner Store Dataset'. On the right, the details for this dataset are displayed, including the NYU logo, the dataset name, a link to the NYU Data Catalog, the creation date, the provider (NYU Health Sciences Library), the author (Brian Elbel), the time period covered (2011), the area covered (New York), and a detailed description of the dataset's content and experimental conditions.

Google Dataset Search

elbel

About

Feedback

20 results found

NYU Bellevue Hospital Experimental Corner Store Dataset
datacatalog.med.nyu.edu
Created May 11, 2017

NYU New York City Low-Income Housing, Neighborhoods and...
datacatalog.med.nyu.edu
Created Jul 21, 2015

NYU Evaluation of the NYC Water Jets Program
datacatalog.med.nyu.edu
Created May 11, 2017

NYU Using Tax Credits to Introduce Supermarkets in Low-Income...
datacatalog.med.nyu.edu
Created May 11, 2017

NYU Impact of New York City Sugar Sweetened Beverage Policy o...
datacatalog.med.nyu.edu
www.da-ra.de
Created May 11, 2017

NYU Bellevue Hospital Experimental Corner Store Dataset

NYU Data Catalog

Dataset created May 11, 2017

Dataset provided by
NYU Health Sciences Library

Authors
Brian Elbel

Time period covered
2011

Area covered
New York

Description
This dataset includes transaction and survey data from a researcher-opened store at Bellevue Hospital, a large hospital in New York City. The store sold a healthier and less-healthy foods and beverages under five conditions: (1) a baseline condition with no special labeling or taxation; (2) highlighting "less healthy" in red capital letters on the price tag; (3) 30% non-itemized tax on less-healthy items; (4) a 30% non-itemized tax on less-healthy items and highlighting "less healthy;" and (5) a 30% itemized, explicit tax on less-healthy items plus highlighting the "less healthy" signage. Each condition was tested for 8-9 days, for a total of 41.5 days. Researchers also collected survey data about what consumers noticed about the various conditions of the experimental store, why they made the purchases they made, and key demographic information.

All Datasets Discoverable Via Google Dataset Search

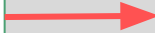
An Extremely Abridged History, Part Two

The Data Catalog Collaboration Project

- Improve data discovery through institutional data catalogs
- Work together to develop, implement, and improve the Data Catalog software

An Extremely Abridged History, Part Two

The Data **Catalog** Collaboration Project

- Improve data discovery through institutional data catalogs
 - Work together to develop, implement, and improve the Data Catalog software
- 

The Data **Discovery** Collaboration Project

- Improve data discovery regardless of tool or platform
- Work towards metadata interoperability between discovery tools
- (Continue developing the Data Catalog software)

Tool Agnostic Data Discovery

[Data Catalog](#) [View Details](#) admin Options

State Inpatient Databases (SID)

[Agency for Healthcare Research and Quality \(AHRQ\)](#)

description

The State Inpatient Databases (SID) are part of the family of databases and software tools developed for the Healthcare Cost and Utilization Project (HCUP). The SID includes inpatient discharge records from community hospitals in that State. The SID files encompass all patients, regardless of payer, providing a unique view of inpatient care in a defined market or State over time.

Record Period Start	Jan. 1, 2011
Record Period End	Jan. 1, 2016
Landing URL	https://www.hcup-us.ahrq.gov/sidoverview.jsp

Keywords Inpatients Community Hospital

Data Storage Requirements: OMB Control # 0915-0276

Data must be stored within the WCM Data Core
Data must be encrypted at rest

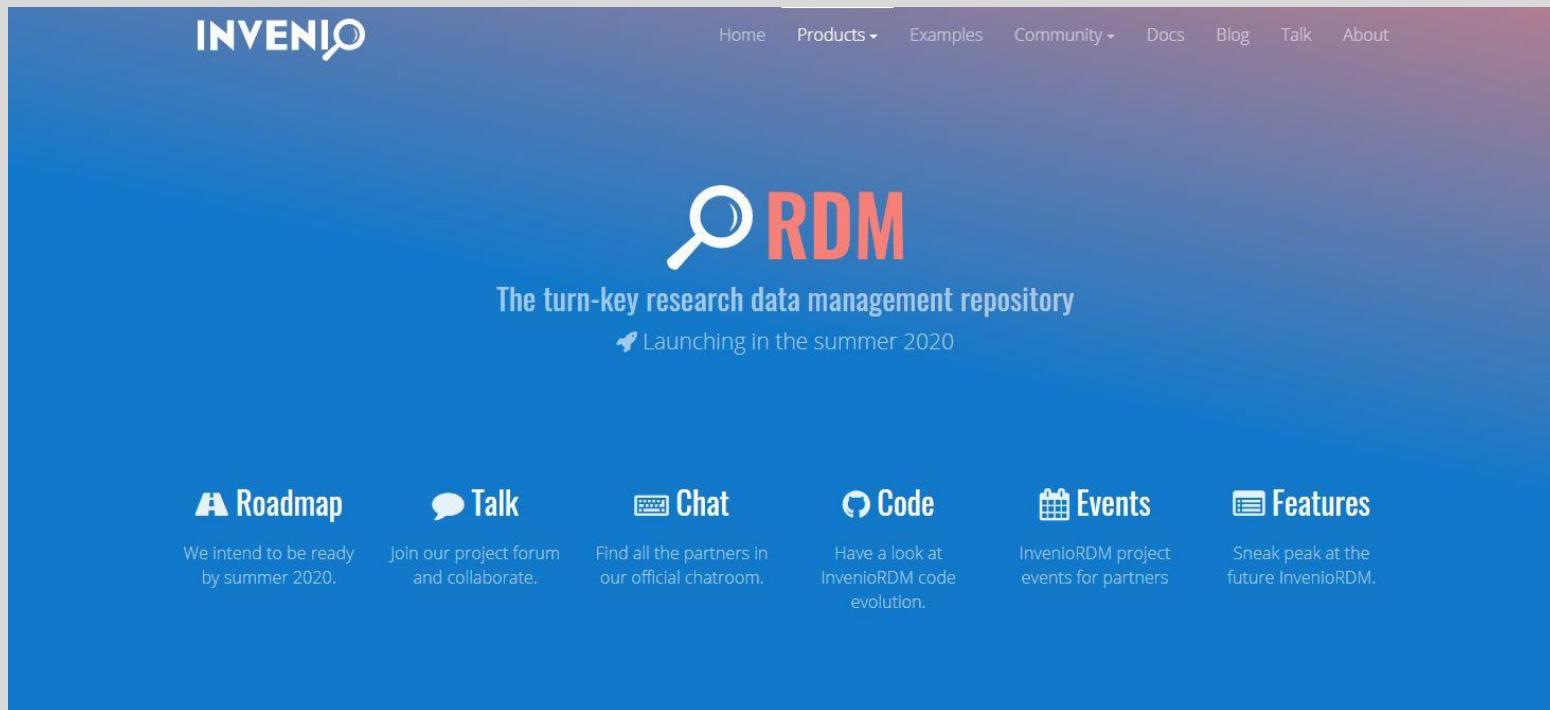
Data Access Conditions: OMB Control # 0915-0276

two-factor authentication

DUAs for State Inpatient Databases (SID)

ID	Title	Data Provider	# Datasets	Scope	Period
OMB Control # 0915-0276	Healthcare Cost and Utilization Project (HCUP)	Agency for Healthcare Research and Quality (AHRQ)	1	User	Nov. 1, 2016 - June 30, 2021

Tool Agnostic Data Discovery



CD2H-Developed RDM Repository

Northwestern


The People

Helenmary Sheridan, MLIS

- Data Services Librarian at the University of Pittsburgh
- In fall 2018, she worked with researcher Charles Horn to catalog software and 3D models created during his research



Describing 3D Models to Improve Discovery

**Pitt Data Catalog**
A project by the Health Sciences Library System

ContactAboutBrowseHome

Find datasets from University of Pittsburgh Researchers

Go!

Include your dataset

3D-Printable Model: Stomach-Vagus Electrophysiology Testing Chamber with Mechanical Stimulus Placement Grid v1.0.0

UID: 22

Author(s): [Horn, Charles C.*](#), [Sciullo, Michael*](#), [Farr, Laura](#)

* Corresponding Author
+ University of Pittsburgh Author

Description

This data package contains 3D printing files (in multiple formats) and instructions for producing a testing chamber.

The gastric vagus supplies critical information to the brain for the control of feeding and aversive signaling, such as nausea and vomiting. This perfusion chamber model (with a grid for testing mechanical stimuli) is designed to test mechanical or chemical stimuli applied to the stomach while recording electrophysiological responses from the vagus, from teased fiber bundles or the whole nerve. The dissected stomach, with attached vagus, of small laboratory animals, including musk shrew (a vomiting species) and mouse (a non-vomiting species),

Access via NIH 3D Print Exchange

Two models in multiple formats, full description, and printing and assembly instructions.
Accession #: 3DPX-003134

Access Restrictions

Free to all

Access Instructions

3D printing files and instructions for printing and assembly available for free download. Licensed under CC BY-SA.

Describing Models at NYU Langone Health

- Working on describing models created by Christopher Collins, PhD at NYU
- Using Helenmary Sheridan's records as a model and building on her expertise

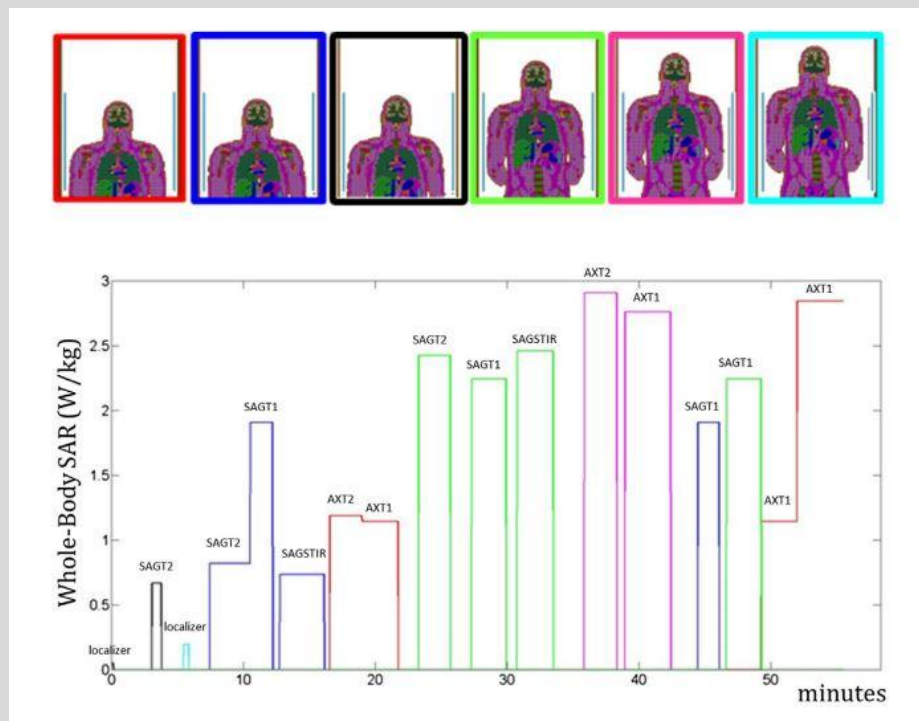


Figure from: Carluccio G, Collins CM. Optimization of the order and spacing of sequences in an MRI exam to reduce the maximum temperature and thermal does. Magn Reson Med. 2019 Mar;81(3):2161-2166. doi: 10.1002/mrm.27503

Carrie Iwema, PhD, MLS, AHIP



Formed the Basic Science Working Group in order to:

- Adapt the metadata to better describe basic science data
- Investigate existing metadata elements to ensure that they work for basic science as well

Read more at: <http://bit.ly/2r6A1OS>

Basic Science Working Group

The image shows a modal window titled "Add New Subject Of Study" with a close button (X) in the top right corner. Inside the modal, there is a label "* Subject of Study" above a text input field. Below this is a label "Species (Strain)" above another text input field. At the bottom of the modal is a "Submit" button. In the background, a larger form is visible with the following elements:

- Text: "Access Res"
- Section Header: "Detailed Dataset Inform"
- Section Header: "Study Type" with two checkboxes: ☐ "Interventional" and ☐ "Observational"
- Text: "Subject of Study" above a text input field, with a link "[Add new](#)" below it.
- Text: "Equipment used to collect/create the dataset" above a text input field, with a link "[Add new](#)" below it.

Suggested changes to:

- Data Type
- Subject of Study
- Equipment
- Software
- Study Type

The Future

Future Growth

- Establish governance structure
- Create an advisory board
- Recruit institutional members who use/develop relevant platforms and tools
- Investigate possibilities for integration of existing tools
- Continue working within our communities to improve data discovery



THE DATA DISCOVERY COLLABORATION PROJECT

**“To enhance discovery of data and other research products
in order to maximize their value”**

What we offer to each other

- Boots-on-the-ground experience with developing metadata to suit researcher needs
 - Experience addressing obstacles to data sharing including researcher apathy and institutional barriers
 - A platform for addressing issues like metadata interoperability
-

Questions?

Nicole Contaxis, MLIS
Lead, NYU Data Catalog
nicole.contaxis@nyulangone.org