DataShare

DataShare:
Lessons from the Field

Julia Kochi and Megan Laurance
University of California, San Francisco Library

E-Science Symposium
April 9, 2014

UCSF LIBRARY
and Center for Knowledge Management

Today I'll Cover…

- Overview of DataShare
- Publicity and Marketing
- Case Studies
- Lessons Learned & Next Steps

Today I'll Cover…

- **Overview of DataShare**
- Publicity and Marketing
- Case Studies
- Lessons Learned & Next Steps

DataShare Overview

- Brainchild of Michael Weiner, M.D.
- Since then:

FAIR ACCESS TO SCIENCE AND TECHNOLOGY RESEARCH ACT (FASTR)

PLOS

UCSF LIBRARY and Center for Knowledge Management

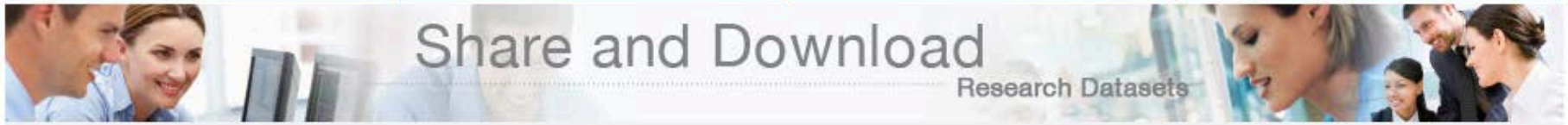
Goal

- Simplify data deposit for UCSF researchers in order to facilitate widespread data sharing
  - User-centered
  - Easy to use interface
  - Self-service
  - Minimal metadata
  - Format and subject agnostic

DataShare

Open data for the global research community

About | Search Data | Share Data (Beta) | My Datasets



Search for Data

Go

or

» Browse all data

Share Data

Make your data count
Get credit for your data
Ensure reproducibility
Promote reuse
Meet funder requirements

» Learn more

Featured Dataset

Progression of white matter degeneration in amyotrophic lateral sclerosis: A diffusion tensor imaging study. White matter damage in frontotemporal dementia and Alzheimer's disease measured by diffusion MRI. Patterns of age-related water diffusion changes in human brain by concordance and discordance analysis.

Researcher Voice

"Making data transparent and available is going to accelerate all of science. It's a relatively inexpensive way to get more value out of all of the work that we do."
-Dr. Michael Weiner, UCSF

CDL | UCSF LIBRARY | Clinical and Translational Science Institute (CTSI)

DataShare

Open data for the global research community

About | Search Data | My Datasets

Steps to Share Your Da

1. Prepare
Gather your data and information before starting. [

2. Describe Data
Create your metadata. [Read More]

3. Upload data
Add your data and metadata to DataShare. [Read More]

4. Get confirmation
Receive your data citation.

Begin

Prompted to log in via campus account

Metadata Fieldname	Description
Title*	Title of the data resource
Data Creator(s)*	Main researcher(s) involved in producing the data, or the authors of the publication, in priority order. (Last Name, First Name)
Description*	An abstract or general description of the data resource
Institution	Name of the institution that supported creation of the resource (DataShare currently pre-populates this field)
Date	Date the data resource was shared (DataShare currently pre-populates this field)
Methods	Any technical or methodological info that helps to understand how the data were generated and/or how they may be properly re-used
Keywords	Any keywords that help identify data domain(s) (DataShare **highly recommends** using a controlled vocabulary, such as MeSH)
Lab / Department	Lab or department responsible for collecting, creating, or otherwise contributing to the development of the dataset
Funder	Agency responsible for funding the data generation
Related Resources	This field may be used to indicate the existence of any other resources that are associated with the data, for example publications or grant numbers (All formats are acceptable, however for publications an identifier is preferred e.g. PMID or DOI)

increases the visibility of underlying research (69% increase in citations for articles associated with shared datasets)

Allows credit for all research outputs, including unpublished work

Enables validation and reproducibility ("An article ... in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship." - D. Donoho)

Enhances global collaborative opportunities

Data re-use maximizes efficiency of available resources

View the DataShare Data Use Agreement

DataShare

Open data for the global research community

Logged in as *420965@ucsf.edu*

» Logout

About | My Datasets

STEP ONE: **Prepare**

STEP TWO **Describ**

Submit

Describe your

* Required Field

* **Dataset Title** e.g., Gu

* **Institution** UCSF

* **Publication Year** e.g, 201

* **Data Type** e.g., Dataset

Data Creator(s) e.g., Weiner, Michael W. + Additional creator

Keyword(s) e.g., Atrophy + Additional keyword

Abstract

Data Type

Use this field to specify the type of file(s) you will be uploading as part of this dataset. DataShare accepts all file formats. Generally speaking, they should fall into one of the following types:

- text
- spreadsheet (csv, tab-delimited, Excel)
- video
- image
- software code
- pdf

If your data files are of a specialized or proprietary nature, be sure to include any pertinent information that would allow the proper viewing and/or usage of the file(s), either as a separate 'Readme' document within the data set folder, or in the 'Technical Description'.

If the dataset consists of multiple files, we recommend, but do not require, that you put all of the files into a single compressed folder (e.g. zip, tar) before uploading it to DataShare.

Upload your datasets to DataShare

Drag & drop files here

…or you can + Add files...

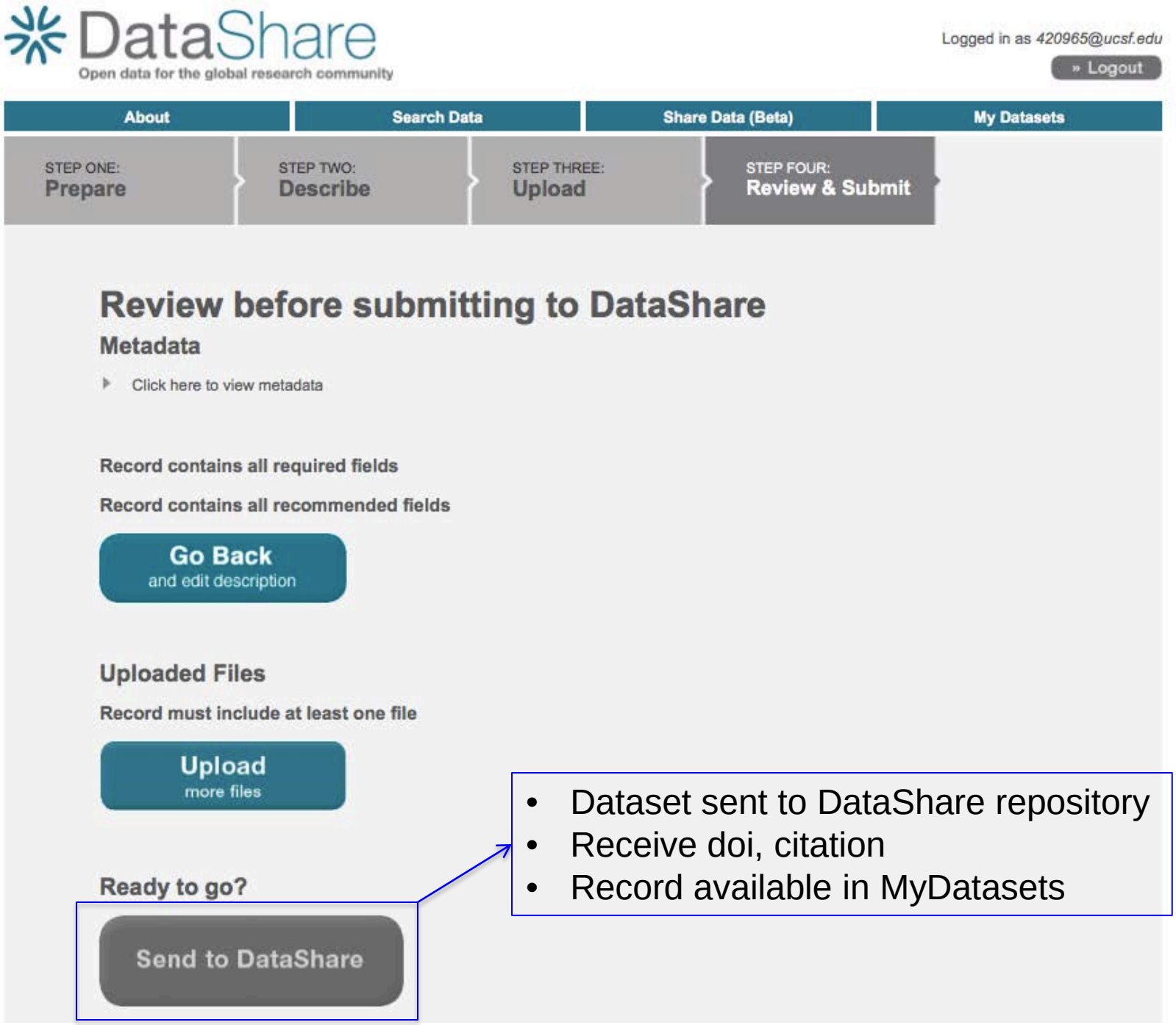
GSM11641..._2_.CEL	5.38 MB	Delete		
GSM11641..._2_.CEL	5.38 MB		Start	Cancel
GSM11641..._2_.CEL	5.38 MB		Start	Cancel

Go Back
and edit description

Review & Submit

Upload FAQ

- Be sure you have read the terms of the DataShare Data Use Agreement
- Ensure that all data relating to human subjects has been properly de-identified
- Ensure data files are the most up-to-date and complete versions
- DataShare accepts all file formats.
- Generally speaking, the data files you upload into DataShare should fall into one of the following types: text, spreadsheet (csv, tab-delimited, Excel), video, image, software code, pdf.
- If the data files are of a specialized or proprietary nature, be sure to include any pertinent information that would allow the proper viewing and/or usage of the file(s) either as a separate 'Readme' document within the data set folder, or in the 'Technical Description'.


DataShare
Open data for the global research community
Logged in as 420965@ucsf.edu
» Logout
About
Search Data
Share Data (Beta)
My Datasets
STEP ONE: Prepare
STEP TWO: Describe
STEP THREE: Upload
STEP FOUR: Review & Submit
Review before submitting to DataShare
Metadata
Click here to view metadata
Record contains all required fields
Record contains all recommended fields
Go Back
and edit description
Uploaded Files
Record must include at least one file
Upload
more files
Ready to go?
Send to DataShare
• Dataset sent to DataShare repository
• Receive doi, citation
• Record available in MyDatasets

DataShare
Open data for the global research community

Logged in as *420965@ucsf.edu*

» Logout

About | Search Data | Share Data (Beta) | My Datasets

My Datasets

Title	Publisher	Publication Year	Action
Re-analysis of micorarray data from rapamycin resistent DLBCL cell lines	University of California, San Francisco	2013	Edit | Delete | Log

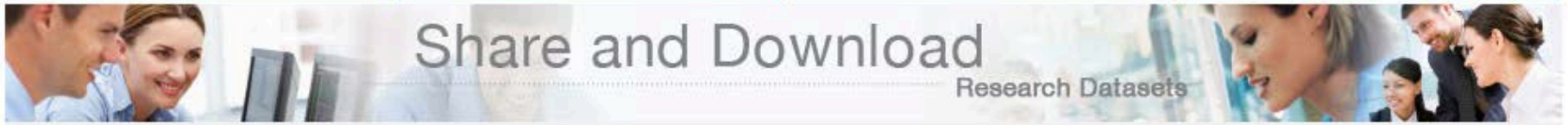
Add new dataset

CDL | UCSF LIBRARY | Clinical and Translational Science Institute (CTSI)

DataShare

Open data for the global research community

About | Search Data | Share Data (Beta) | My Datasets



Search for Data

Go

or

» Browse all data

Share Data

Make your data count
Get credit for your data
Ensure reproducibility
Promote reuse
Meet funder requirements

» Learn more

Featured Dataset

Progression of white matter degeneration in amyotrophic lateral sclerosis: A diffusion tensor imaging study. White matter damage in frontotemporal dementia and Alzheimer's disease measured by diffusion MRI. Patterns of age-related water diffusion changes in human brain by concordance and discordance analysis.

Researcher Voice

"Making data transparent and available is going to accelerate all of science. It's a relatively inexpensive way to get more value out of all of the work that we do."
-Dr. Michael Weiner, UCSF

CDL | UCSF LIBRARY | Clinical and Translational Science Institute (CTSI)

DataShare

Open data for the global research community

About | Search Data | Share Data (Beta) | My Datasets

Select a Dataset...

Contributor

- UCSF Center for Imaging of Neurodegenerative Diseases (12)
- UCSF Bixby Center for Global Reproductive Health (1)

Author

- Weiner, Michael W. (12)
- Schuff, Norbert (7)
- Cardenas, Valerie (3)
- Zhang, Yu (3)
- Chao, Linda (2)

more

Keyword

less

- (1)
- A Beta (1)
- Adult (11)
- Aged (6)

Sorted by: relevance Go!

Associations between vascular risk factors, carotid atherosclerosis and cortical volume and thickness in older adults

by Cardenas, Valerie | Reed, Bruce | Chao, Linda | Chui, Helena | Sanossian, Nerses | DeCarli, Charles | Mack, Wendy | Kramer, Joel | Hodis, Howard | Yan, Mingzhu | Buonocore, Michael | Carmichael, Owen | Jagust, William J. | Weiner, Michael W.
at UCSF Center for Imaging of Neurodegenerative Diseases, University of California, San Francisco
Show abstract

CSF Biomarker and PIB-PET Derived Beta-Amyloid Signature Predicts Metabolic, Grey Matter and Cognitive Changes in Non-Demented Subjects

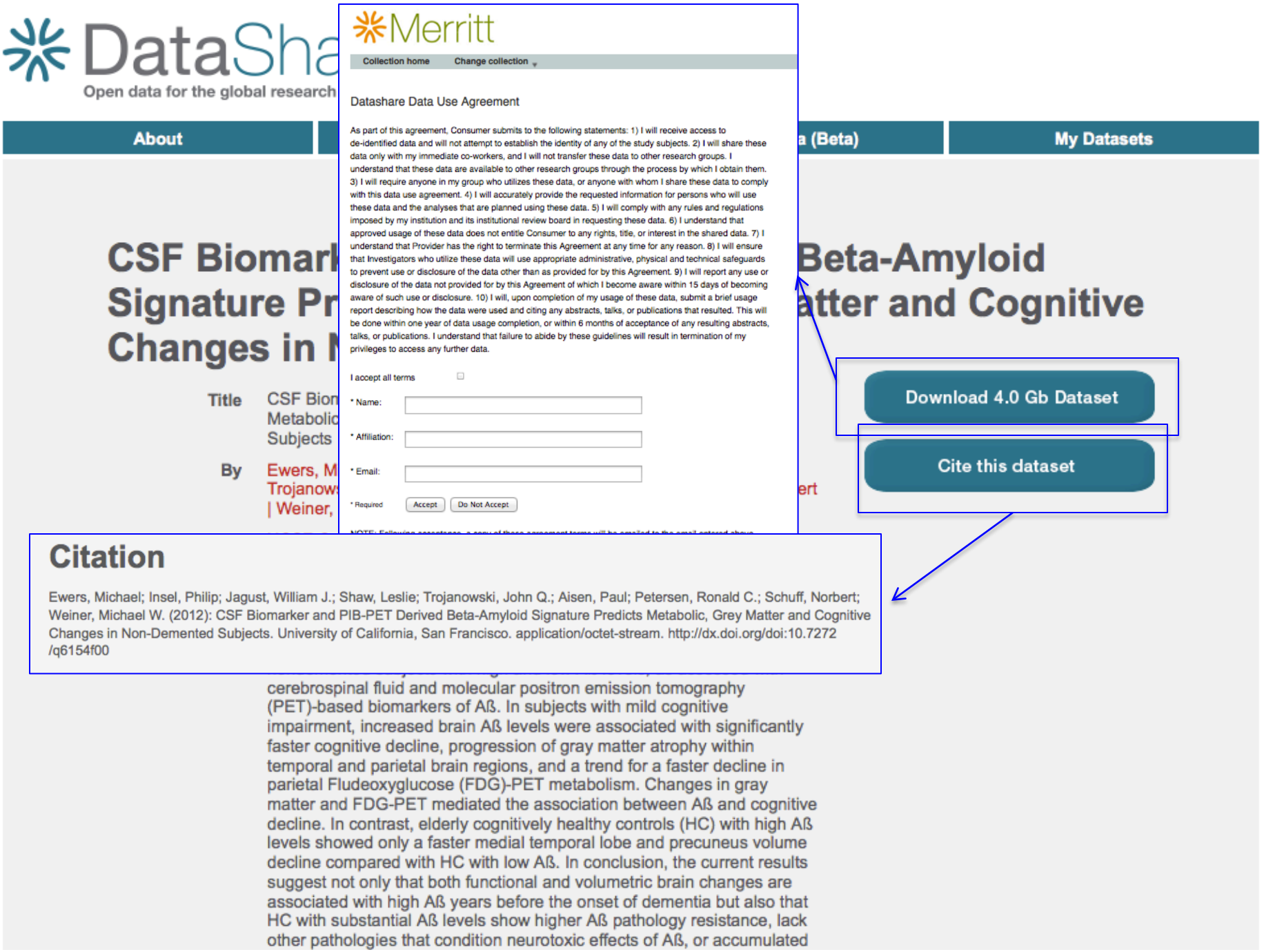
by Ewers, Michael | Insel, Philip | Jagust, William J. | Shaw, Leslie | Trojanowski, John Q. | Aisen, Paul | Petersen, Ronald C. | Schuff, Norbert | Weiner, Michael W.
at UCSF Center for Imaging of Neurodegenerative Diseases, University of California, San Francisco
Show abstract

Frontotemporal Lobar Degeneration (FTLD)

by Weiner, Michael W.
at UCSF Center for Imaging of Neurodegenerative Diseases, University of California, San Francisco
Show abstract

Gulf War Illness

by Weiner, Michael W.
at UCSF Center for Imaging of Neurodegenerative Diseases, University of California, San


DataSha
Open data for the global research
About
(Beta)
My Datasets
Merritt
Collection home
Change collection
Datashare Data Use Agreement
As part of this agreement, Consumer submits to the following statements: 1) I will receive access to de-identified data and will not attempt to establish the identity of any of the study subjects. 2) I will share these data only with my immediate co-workers, and I will not transfer these data to other research groups. I understand that these data are available to other research groups through the process by which I obtain them. 3) I will require anyone in my group who utilizes these data, or anyone with whom I share these data to comply with this data use agreement. 4) I will accurately provide the requested information for persons who will use these data and the analyses that are planned using these data. 5) I will comply with any rules and regulations imposed by my institution and its institutional review board in requesting these data. 6) I understand that approved usage of these data does not entitle Consumer to any rights, title, or interest in the shared data. 7) I understand that Provider has the right to terminate this Agreement at any time for any reason. 8) I will ensure that Investigators who utilize these data will use appropriate administrative, physical and technical safeguards to prevent use or disclosure of the data other than as provided for by this Agreement. 9) I will report any use or disclosure of the data not provided for by this Agreement of which I become aware within 15 days of becoming aware of such use or disclosure. 10) I will, upon completion of my usage of these data, submit a brief usage report describing how the data were used and citing any abstracts, talks, or publications that resulted. This will be done within one year of data usage completion, or within 6 months of acceptance of any resulting abstracts, talks, or publications. I understand that failure to abide by these guidelines will result in termination of my privileges to access any further data.
I accept all terms
* Name:
* Affiliation:
* Email:
* Required
Accept
Do Not Accept
CSF Biomark
Signature Pr
Changes in N
Beta-Amyloid
atter and Cognitive
Title
CSF Biom
Metabolic
Subjects
By
Ewers, M
Trojanow
| Weiner,
ert
Download 4.0 Gb Dataset
Cite this dataset
Citation
Ewers, Michael; Insel, Philip; Jagust, William J.; Shaw, Leslie; Trojanowski, John Q.; Aisen, Paul; Petersen, Ronald C.; Schuff, Norbert; Weiner, Michael W. (2012): CSF Biomarker and PIB-PET Derived Beta-Amyloid Signature Predicts Metabolic, Grey Matter and Cognitive Changes in Non-Demented Subjects. University of California, San Francisco. application/octet-stream. http://dx.doi.org/doi:10.7272/q6154f00
cerebrospinal fluid and molecular positron emission tomography (PET)-based biomarkers of Aß. In subjects with mild cognitive impairment, increased brain Aß levels were associated with significantly faster cognitive decline, progression of gray matter atrophy within temporal and parietal brain regions, and a trend for a faster decline in parietal Fludeoxyglucose (FDG)-PET metabolism. Changes in gray matter and FDG-PET mediated the association between Aß and cognitive decline. In contrast, elderly cognitively healthy controls (HC) with high Aß levels showed only a faster medial temporal lobe and precuneus volume decline compared with HC with low Aß. In conclusion, the current results suggest not only that both functional and volumetric brain changes are associated with high Aß years before the onset of dementia but also that HC with substantial Aß levels show higher Aß pathology resistance, lack other pathologies that condition neurotoxic effects of Aß, or accumulated

Today I'll Cover…

- Overview of DataShare
- **Publicity and Marketing**
- Case Studies
- Lessons Learned & Next Steps

Publicity and Marketing

- Researchers with funder/publisher requirements
- Data Managers
- Lab Managers
- Recipients of OA funding

- OA Week events
- Group presentations
- Campus-wide tech fair
- Integrated into postdoc and graduate level courses

UCSF LIBRARY
and Center for Knowledge Management

Today I'll Cover…

- Overview of DataShare
- Publicity and Marketing
- **Case Studies**
- Lessons Learned & Next Steps

Case Study 1

- *"We're done with this data. We've picked all the fruit."* - *Data Manager, Women's Health Research Group*
  - NIH study concluded, findings published
  - Requests from other researchers
  - No obvious home for data (e.g., subject repository)

Case Study 2

- *I want to share my data, but my current method is costly.* - *Professor, Dept. of Biochemistry and Biophysics*
  - Repository currently using too cumbersome
  - Actual cost vs time cost
  - Dataset size
  - Long-term storage

Case Study 3

- *I need something for a grant I'm writing.*
  - *Divisional Analyst, OBGYN & Associate Professor, Dept. of Medicine*
  - Immediate need to fulfill grant proposal

Other Use Cases

- Outside collaborators
- HIPPA compliance
- Continuously updated dataset

Today I'll Cover…

- Overview of
- Publicity and Marketing
- Case Studies
- **Lessons Learned & Next Steps**

Lessons Learned

- It’s hard to get noticed
- It’s even harder to break existing patterns of data sharing
- User feedback is key
- Insert at point of pain/need

Next Steps

- Continue usability testing and user interviews
- More outreach and marketing
- Develop stronger partnerships
- Develop a deeper data management toolkit
- Participate in further development of DataShare at a UC-wide level

Acknowledgments

Mini Kahlon
Angela Rizk-Jackson
Anirvan Chatterjee

CTSI at UCSF

NCATS – NIH Grant # UL1 TR000004

Michael Weiner

UCSF

Karen Butter
Julia Kochi
Megan Laurance
Geoffrey Boushey

UCSF LIBRARY and Center for Knowledge Management

Patricia Cruse
Carly Strasser
Perry Willett
Stephen Abrams
Marisa Strong
Eric Hetzner & Abhishek Salve

University of California CDL California Digital Library

Contact Info

datashare.ucsf.edu

Julia Kochi: julia.kochi@ucsf.edu

Megan Laurance: megan.laurance@ucsf.edu

UCSF LIBRARY
and Center for Knowledge Management