

# Image across Domains, Experiments, Algorithms and Learning (IDEAL)

Dani Ushizima<sup>1,2</sup>

<sup>1</sup>Staff Scientist, <sup>1</sup>CAMERA Image Processing Team Lead, <sup>2</sup>Data Scientist Fellow

<sup>1</sup>DAV, CRD, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

<sup>2</sup>Berkeley Institute for Data Sciences, UC Berkeley, CA, USA



# Image across Domains, Experiments, Algorithms and Learning (IDEAL) – P.I. Ushizima

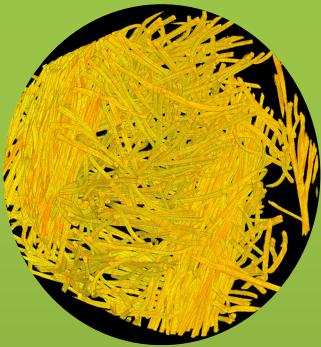
DOE ASCR Early Career Research Program 2015-2020

## Mission:

- Aggregate value to data collected at DOE facilities;
- Help scientists discover relevant, but hidden information from digital images;
- Focus on DOE research across domains reliant on experimental observation data;
- Recognize patterns through machine learning and multiscale analyses, which is delivered as user-friendly software;
- Multidisciplinary work to accelerate analytics, reducing time between experiments, and opening more opportunities for more users of the imaging facilities.

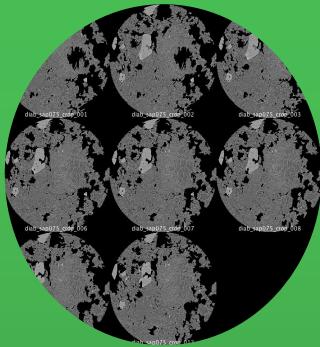


# Image Across Domains



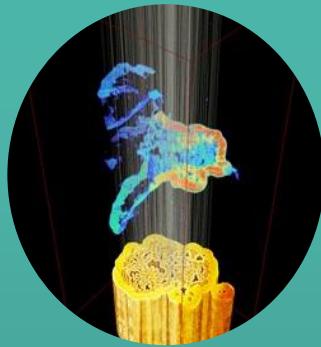
## Specimens

- Materials, composites, polymers, other compounds and biological samples.



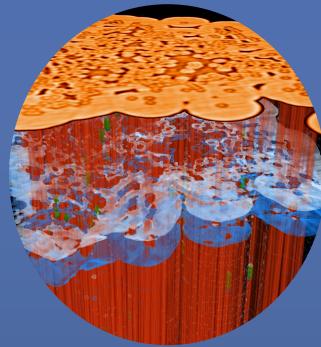
## Formats

- Tiff, jpeg, hdf5, bioformats, feature vectors, multi-resolution image pyramids, e.g Laplacian, Wavelets.



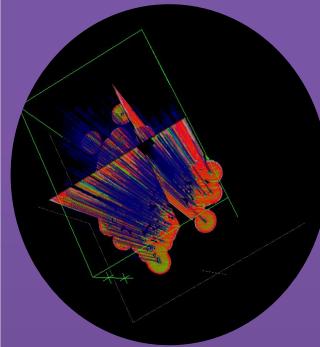
## Data Analysis

- Morphometry;
- Spectral content;
- Multimodal sources;
- Templates.



## Data Inference

- Clustering;
- Classification;
- Deep learning;
- Decision-making
- Visualization.

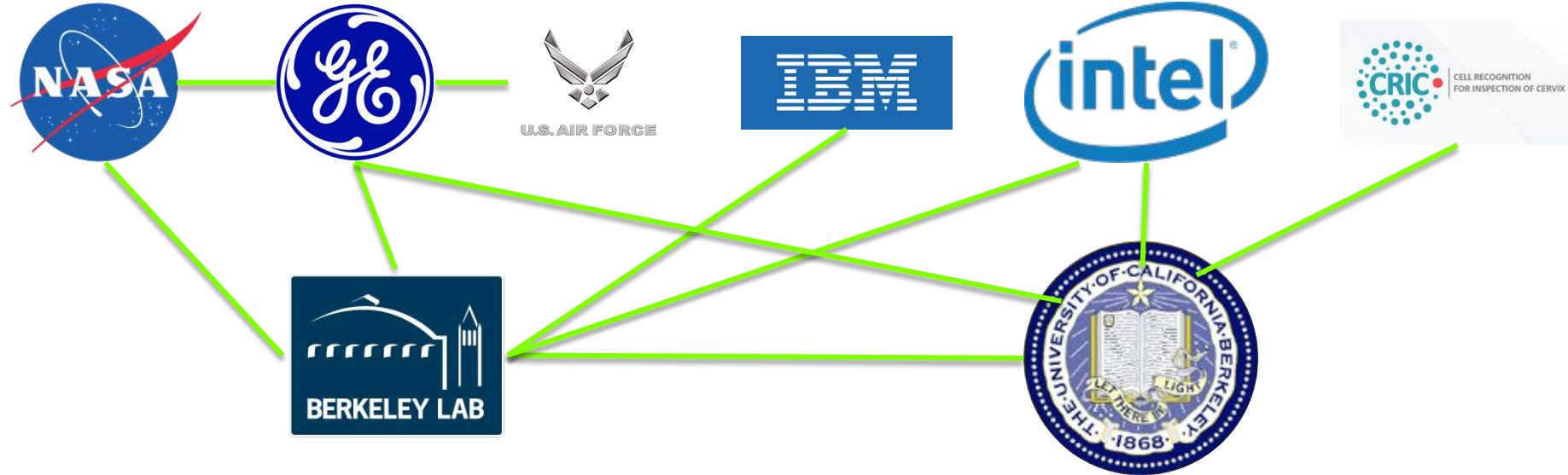


## Reproducible research

- Data repositories;
- Software repositories;
- Motifs;
- Searchable datasets.



IDEAL - Dani Ushizima



GORDON AND BETTY  
**MOORE**  
FOUNDATION



Science  
Without Borders



Microsoft

FAPEMIG

SEBRAE

- Collaborations among national labs, academia and industry

# USE-CASES

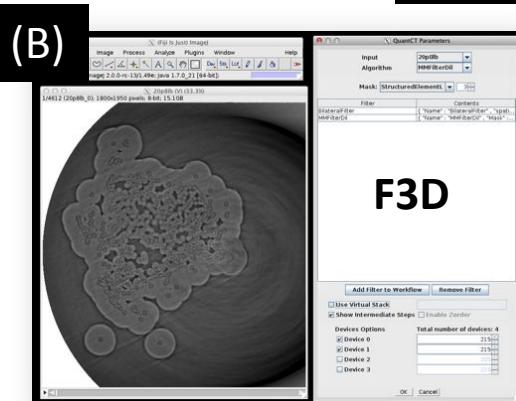
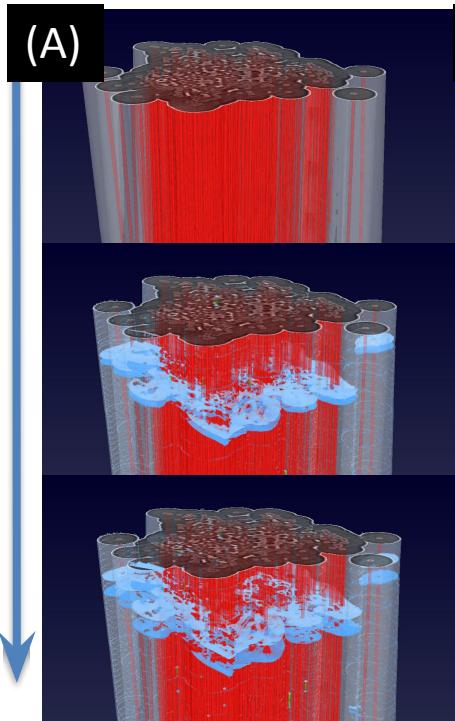
# Quantification of micro-structural damage



## Scientific Achievement

- Deeper **understanding** of material deformations and endurance;
- New algorithms/computational methods to **recognize patterns** repeating **across samples**;
- **Quantify** microCT data for micro-structural damage to safely deploy new materials.

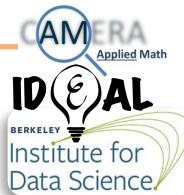
ALS



(A) Rendering of samples under deformation after image processing and analysis to detect the matrix (gray) and fibers (red) of a single tow at different applied loads, leading to microfractures (blue); (B) Deployed tool: F3D.

## Significance and Impact

- Support ALS data characterization;
- Provide **software** that works to multiple problems, e.g. ceramics, cement, textile;
- Fast characterization of materials properties: F3D is **17x faster** on a experiment (~60GB);



## Research Details

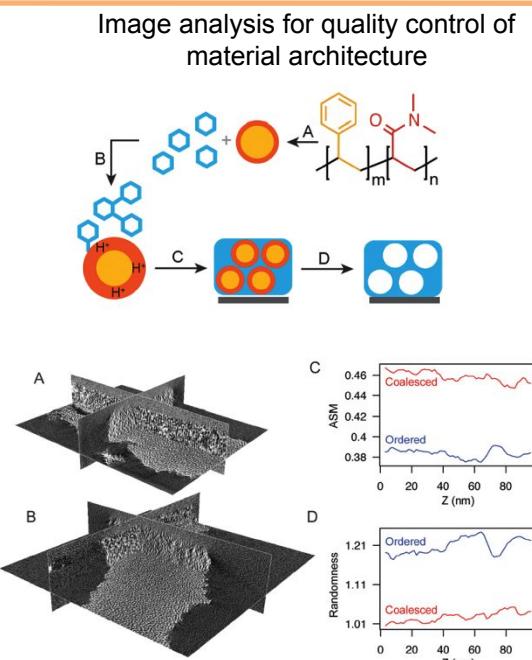
- Automatic identification of **microdamage** with fast-filtering (data streaming) through GPUs;
- Analysis of ceramic-fiber **composite** used in jets;
- ALS MicroCT used to image **microfractures** and matrix cracks under increasing strain;
- Efficient analysis = **near real-time**;
- Reduced turnaround: days to min.

# Quality control during assembly of films



## Scientific Achievement

- Improve film manufacturing by quantifying **pore structure** evolution;
- New methodology to describe **material architecture** enhancement;
- **Analysis** of thin films by using scanning transmission electron microscopy tomogram.



## Significance and impact

- High-quality film to be used by Intel;
- Identified ideal fabrication **conditions** by measuring porosimetry;
- **New tools** catalyzed the collaboration among Intel, LBL NCEM and Organic and Macromolecular Synthesis at the Molecular Foundry, and SLAC.

NCEM  
Molecular Foundry



## Research details

- Reported **lowest ever dielectric constants** for PMO matrix material, used in **microelectronics**;
- Delivered **analysis method**, second-order statistics of image intensity variations, to measure film roughness;
- New tools adapted to 3D stacks for **NCEM instruments**;
- **Upcoming developments:** nanoframework analysis using **machine learning** (with T. Williams on Chemical Comm. 2017).

# Light for studying matter, CNN for pattern



## Scientific Achievement

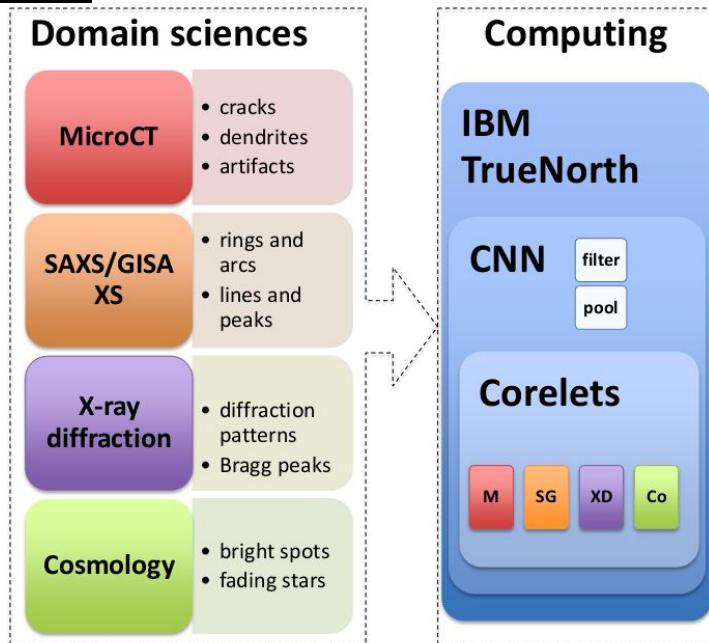
- Devise algorithms that **generalize across** scientific domains (Fig. A);
- Recognize patterns from data **without** feature design;
- Convolutional neural networks (CNN) runs on low-energy chips.

## Significance and impact

- MicroCT and Xray diffraction prototypes with **accuracy** above **98%** (Fig. B and C);
- Deployed **recognition** pipelines that run in a single IBM TN chip.



(A)

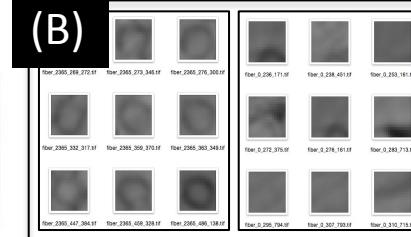


ASCR/BES/BER

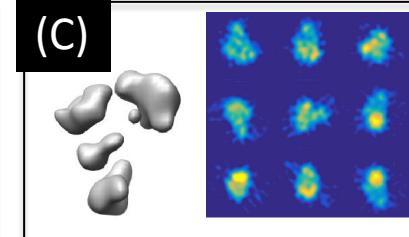
## Research details

- Exploration of experimental **observational** and **simulation** data;
- Ability to use previous manually created data to **automate** analysis;
- Tested algorithms on CPU, GPU and **neuromorphic** architectures.

(B)



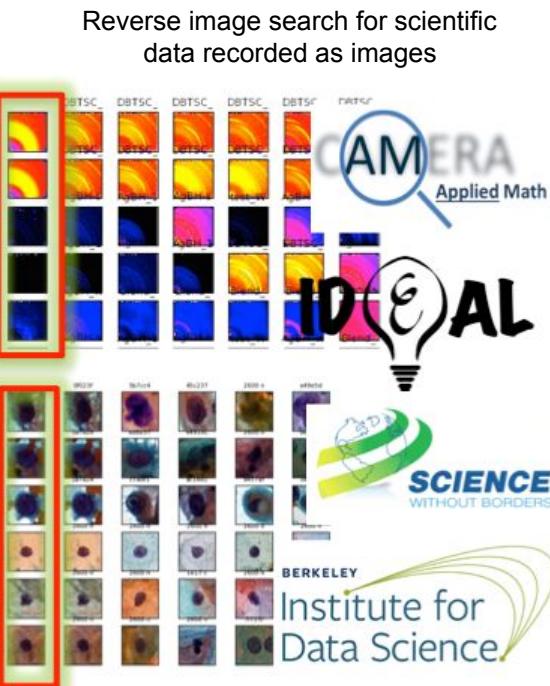
(C)



# Recommendation System for scientific image

## Scientific Achievement

- New **visual search engine**: pyCBIR, for scientific image retrieval based on pictorial similarity;
- Tool-chain capable of retrieving **relevant** images using datasets across science domains;
- **Convolutional Neural Nets**.



Computer, Math,  
Pathology, Physics

## Significance and impact

- **Real-time** image retrieval using compact data representation;
- **Enable** investigation of **abstract patterns**, by leveraging historical data, gathered by domain experts at a high cost;
- **Improve** collaboration among researchers across scientific communities.

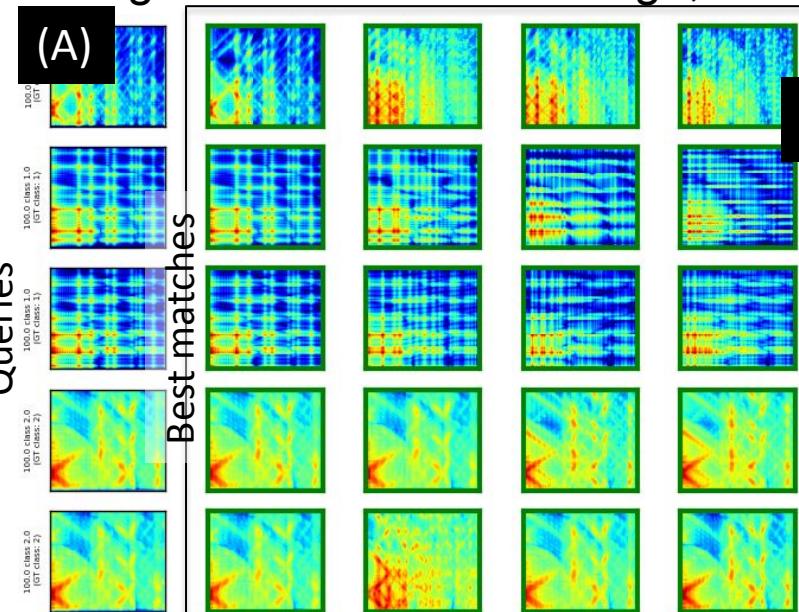
## Research details

- Developed **CNN**-based tools for pattern recognition using optimized libraries, such TensorFlow, cuDNN, cuFFT;
- Created **quantitative analytics** to recover data and respective **confidence** associated to top-matches;
- Tested pyCBIR with images from **different spatial scales**;
- **Upcoming developments:** promote guided data explorations and new discoveries among scientists

# Deep learning for nanostructure from X-ray diffraction

## Scientific Achievement

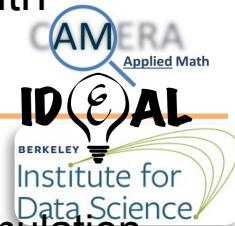
- Search and ranking materials using X-ray diffraction data (Fig. A);
- Recognize crystal structure from images **without** feature design;



ALS/CRD

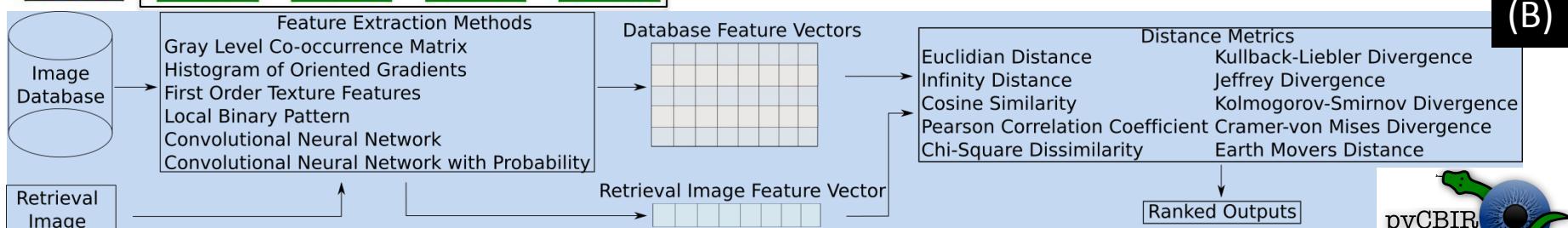
## Significance and impact

- Devise algorithms to **bridge** the gap between theoretical models and experimental observational data;
- X-ray diffraction prototypes with **accuracy** above **98%**;



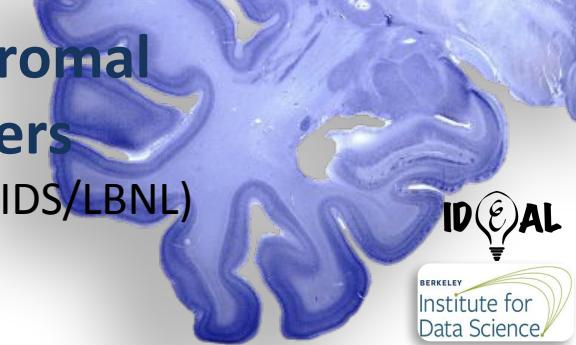
## Research details

- Exploration of ALS data, including simulation performed @NERSC using machine learning (ML) prototype (Fig.B) from CAMERA;
- Test ability to categorize millions of GISAXS patterns without manual interaction;
- Core codes: HipGISAXS simulation @NERSC, ML TensorFlow @CAMERA GPU test-bed.

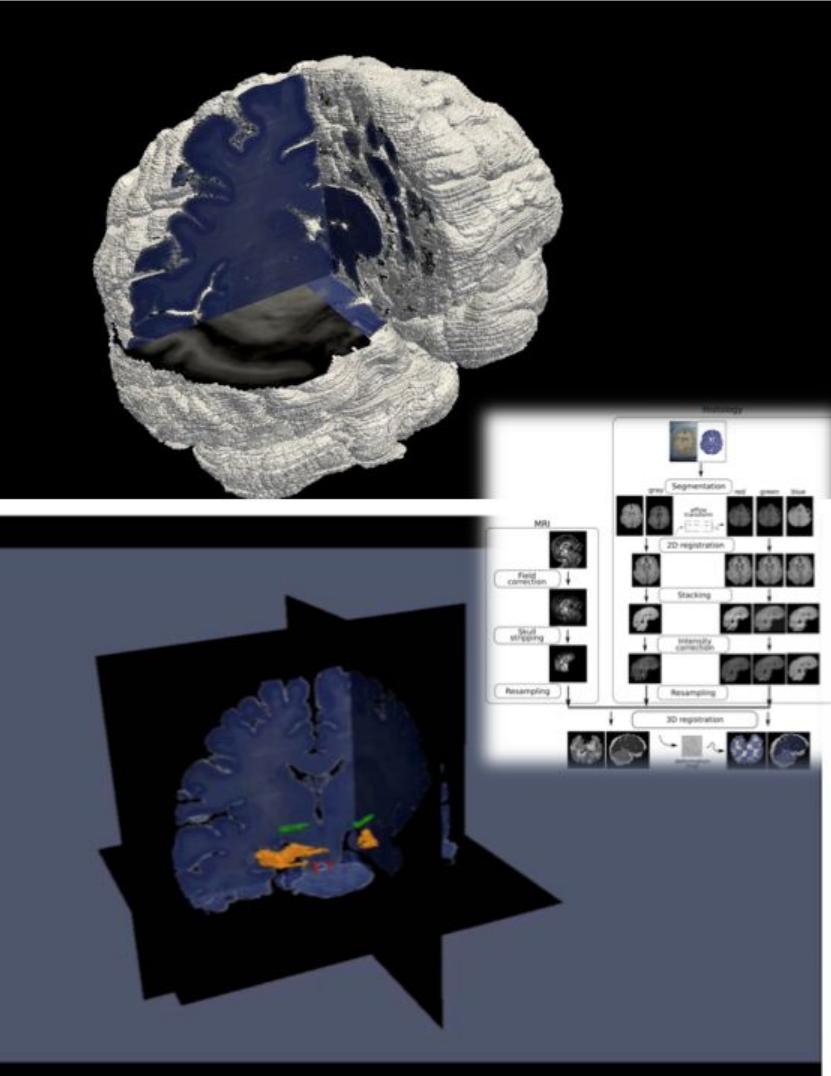


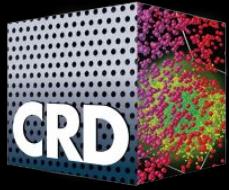
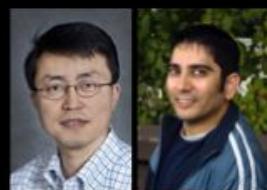
# Pathway toward diagnosing and monitoring prodromal Alzheimer's Disease using novel imaging biomarkers

Alegro PhD, Grinberg, M.D, Ph.D (UCSF) and Dani Ushizima, Ph.D (BIDS/LBNL)



- Problem:
  - Alzheimer: 5.4 million Americans, over 600K in CA, but no cure available yet;
- Approach:
  - UCSF: (MRI)-based biomarkers and potential to detect and track Alzheimer's progression in early stages;
  - LBNL: Computational pipeline for **registration** of whole brain histological images to corresponding MRI data;
- Impact:
  - Software tools to improve clinical MRI resolution with histology and multimodal imaging;
  - Support development of new drugs that target individuals more precisely.



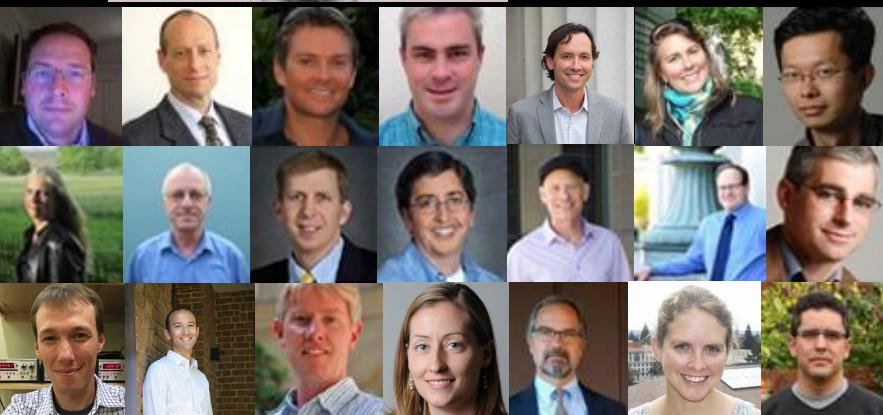


Data Analytics and Visualization  
Group





# BERKELEY INSTITUTE FOR DATA SCIENCE



Berkeley  
UNIVERSITY OF CALIFORNIA

# Disclaimer

- This presentation was prepared with basis on work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.