Advanced Cyberinfrastructure

to enable

Multi-scale and Multi-site Brain Research

Mark Ellisman

Professor of Neurosciences and Bioengineering,
University of California San Diego,
and Director of the Center for Research in Biological Systems

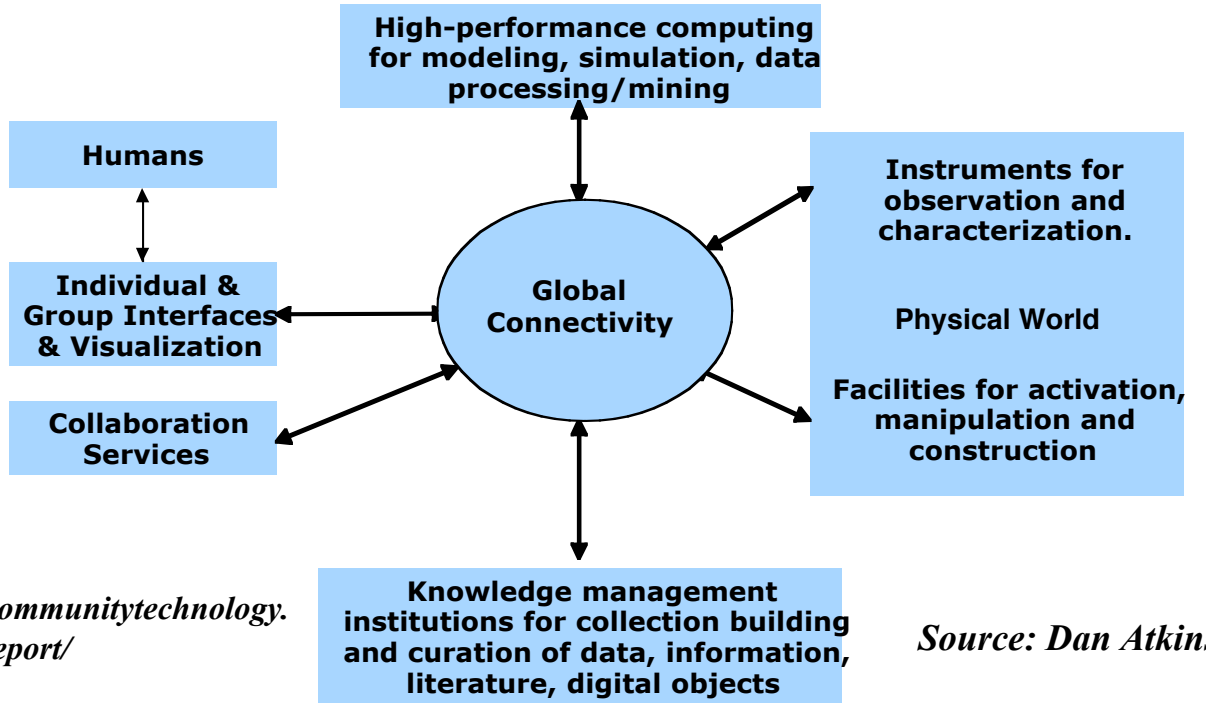
"CYBERINFRASTRUCTURE"

What is a Grid? What do we mean?

- **GRID technologies bring remote resources together**

A broad, systemic, strategic conceptualization

Components of Cyberinfrastructure
(Grid)-enabled science & engineering



http://www.communitytechnology.org/nsf_ci_report/

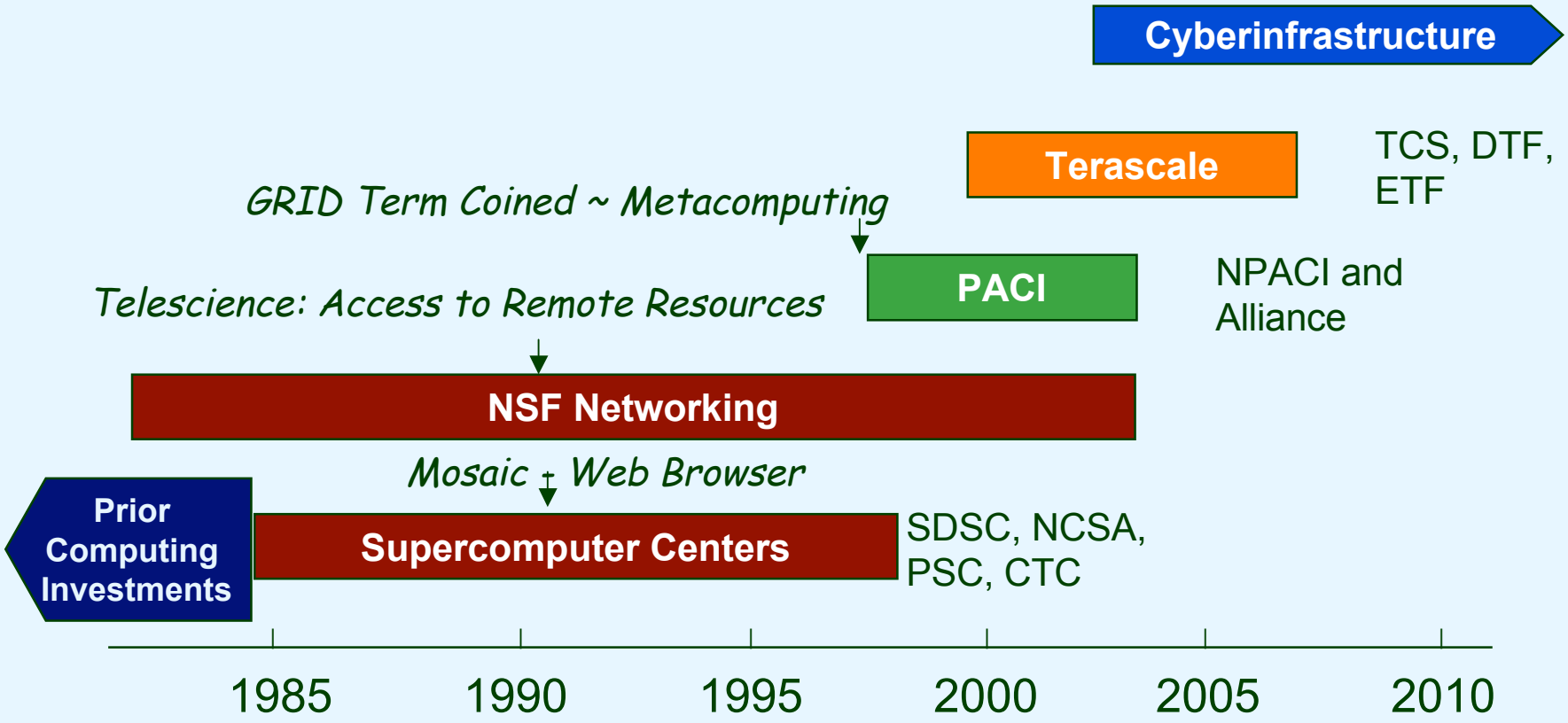
Source: Dan Atkins

Grid implies global (international) system for collaboration

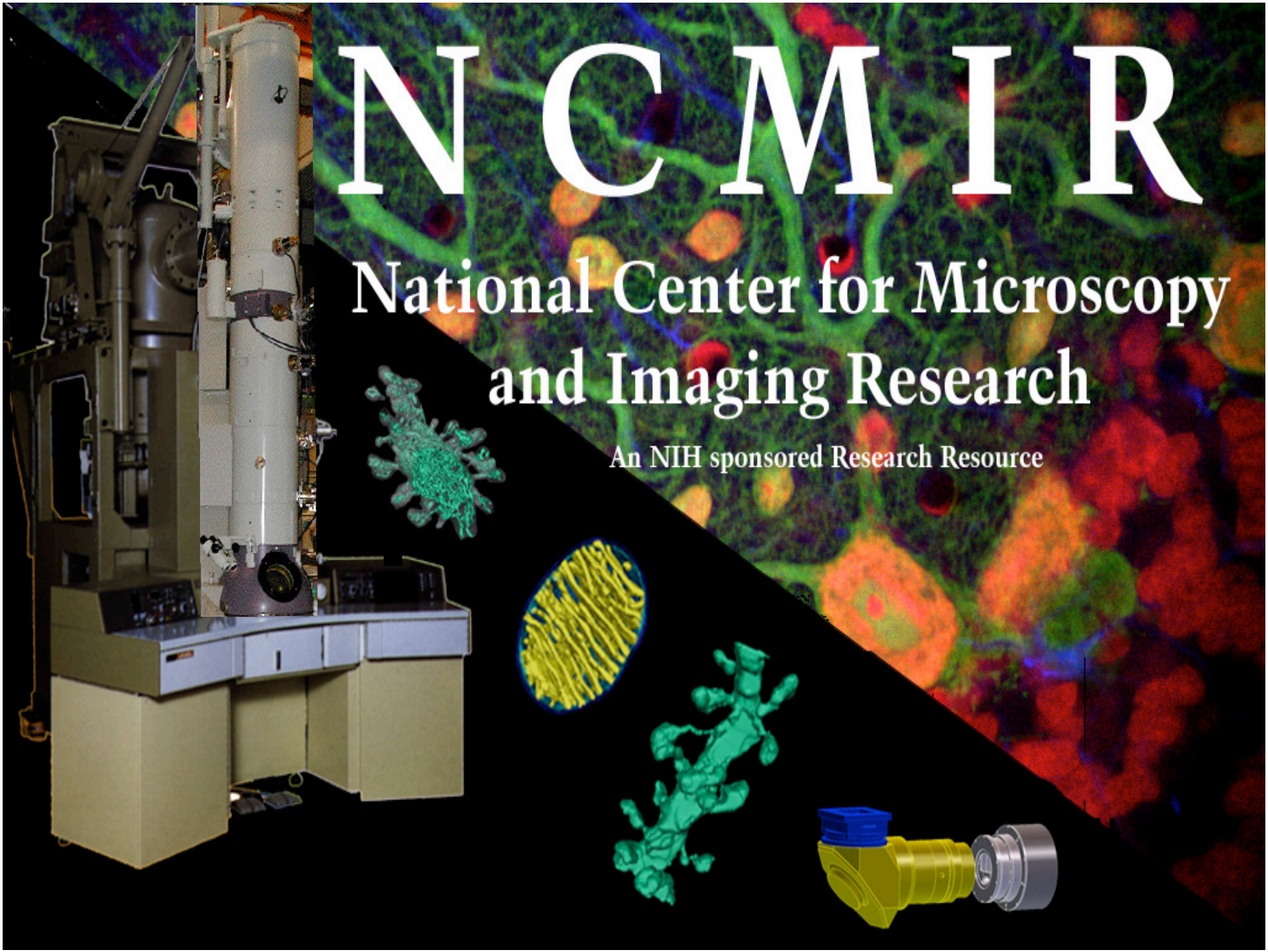
Evolution of the Computational Infrastructure

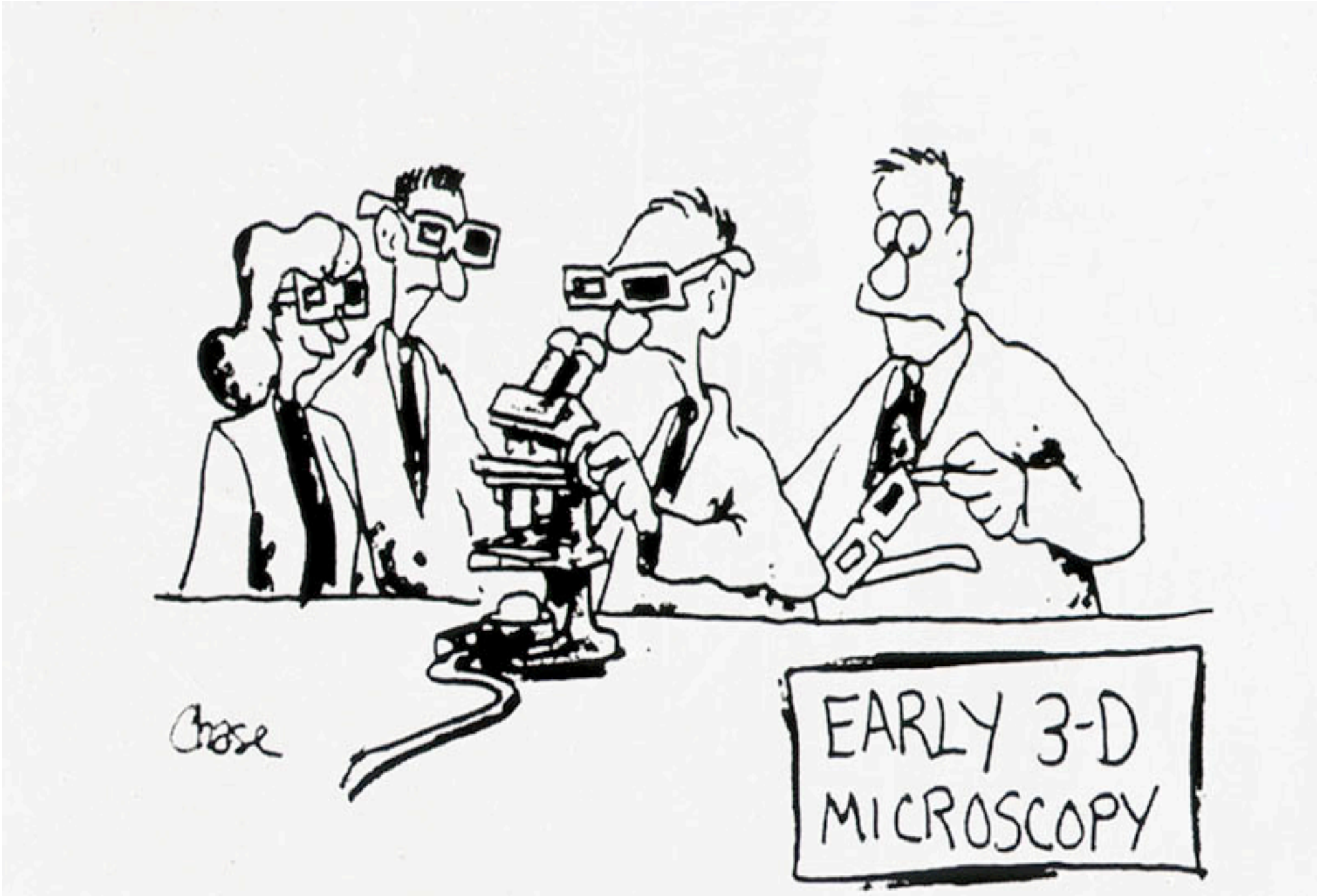
Investments in the US *Source: Dr. Deborah Crawford*

Chair, NSF ***CyberInfrastructure*** *Working Group (CIWG)*



A timeline from the Computational Infrastructure Division of the US National Science Foundation


NCMIR
National Center for Microscopy
and Imaging Research
An NIH sponsored Research Resource


Crose
EARLY 3-D
MICROSCOPY

High Energy Electron MICROSCOPY

Ultra High Voltage EM @ Korea Basic Sciences Inst..

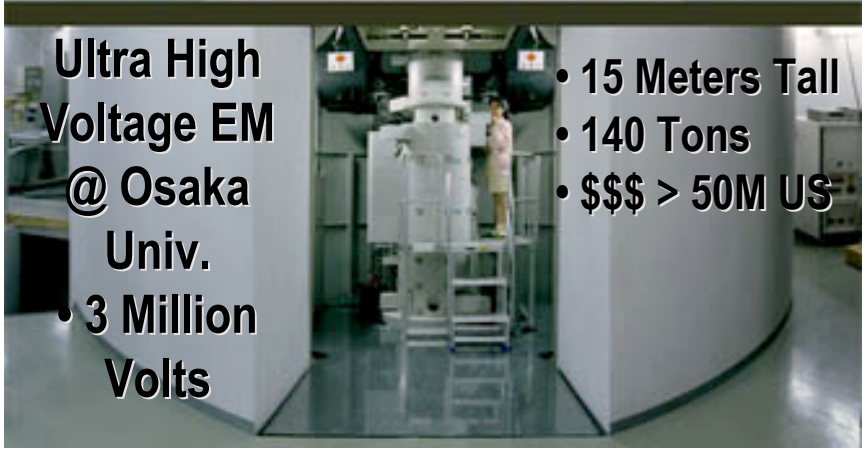
- 1.25 Million Electron Volts

KOREA

JAPAN

Ultra High Voltage EM @ Osaka Univ.

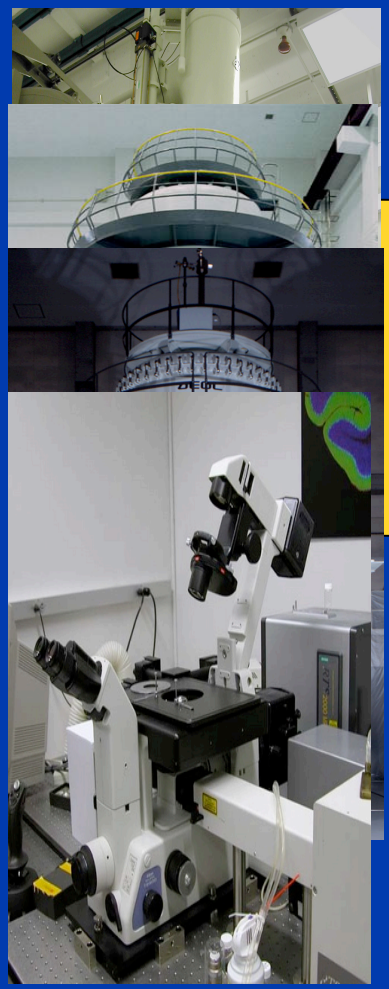
- 3 Million Volts



- 15 Meters Tall
- 140 Tons
- $$$ > 50M US

TELEMICROSCOPY & CYBERINFRASTRUCTURE

REMOTE ACCESS FOR DATA ACQUISITION, GRID COMPUTING AND DISTRIBUTED DATABASES



Server Software

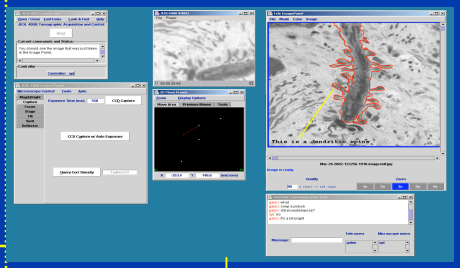
Instrument Interface

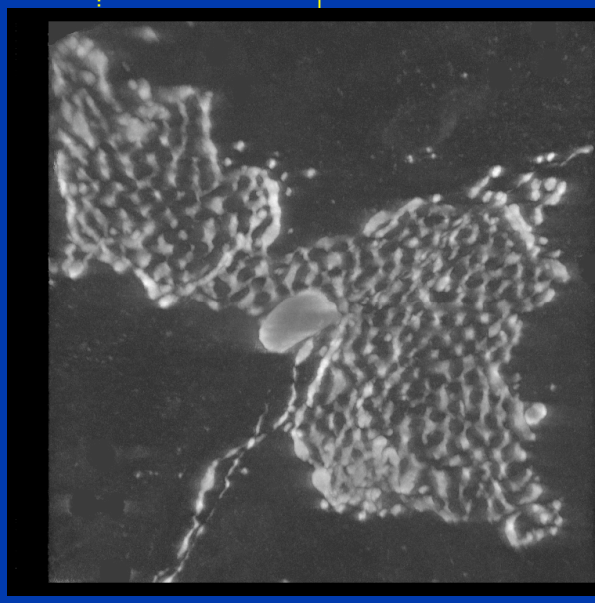
Session Manager Service

Grid Services

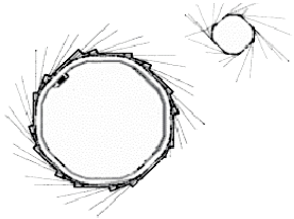
Data Storage | Computation

Web Browser



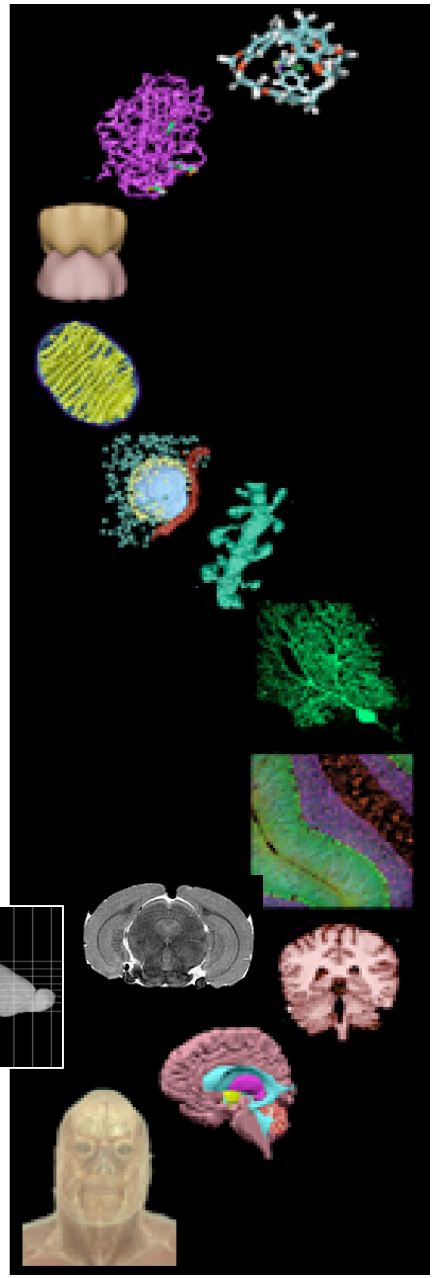


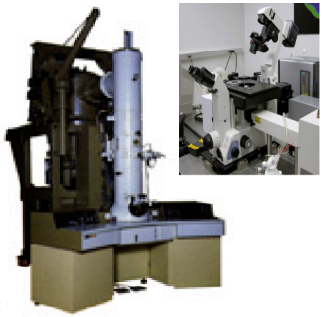
Synchrotrons, Microscopes and MRI's: Tools for the Nano, "Meso" and Macro Scales of Biological Systems



Synchrotrons

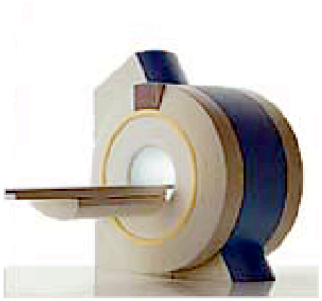
Molecules



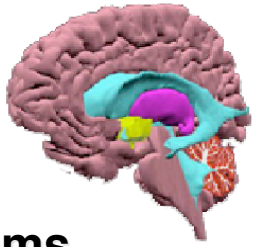


Microscopes

Macromolecular Complexes, Organelles, Cells



Magnetic Resonance Imagers



Organs, Organ Systems, Organisms

The Login Page: The Entrance into the Portal

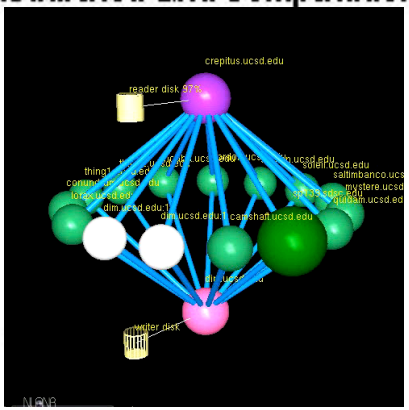
Single login grants authenticated access to all applications, resources, and services

telescience

for advanced tomography applications

Remotely Control Instruments

Distributed Grid Computation



Distribute Data to Storage Resources

Username

Password

Login

What is Telescience?

Learn more about the Telescience Portal

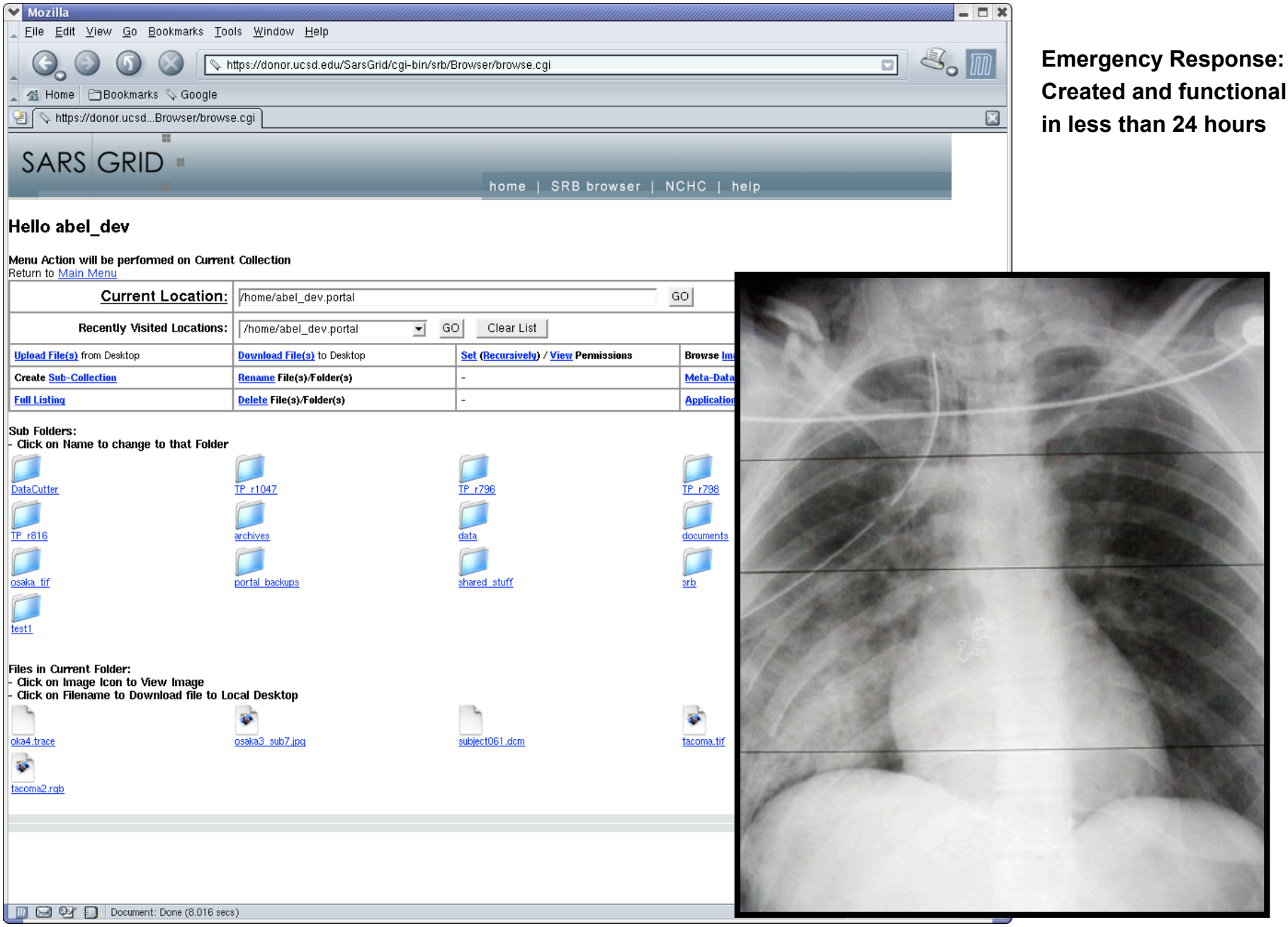
Need a Telescience Account?

NCMIR NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH at San Diego, an NIH supported research resource center

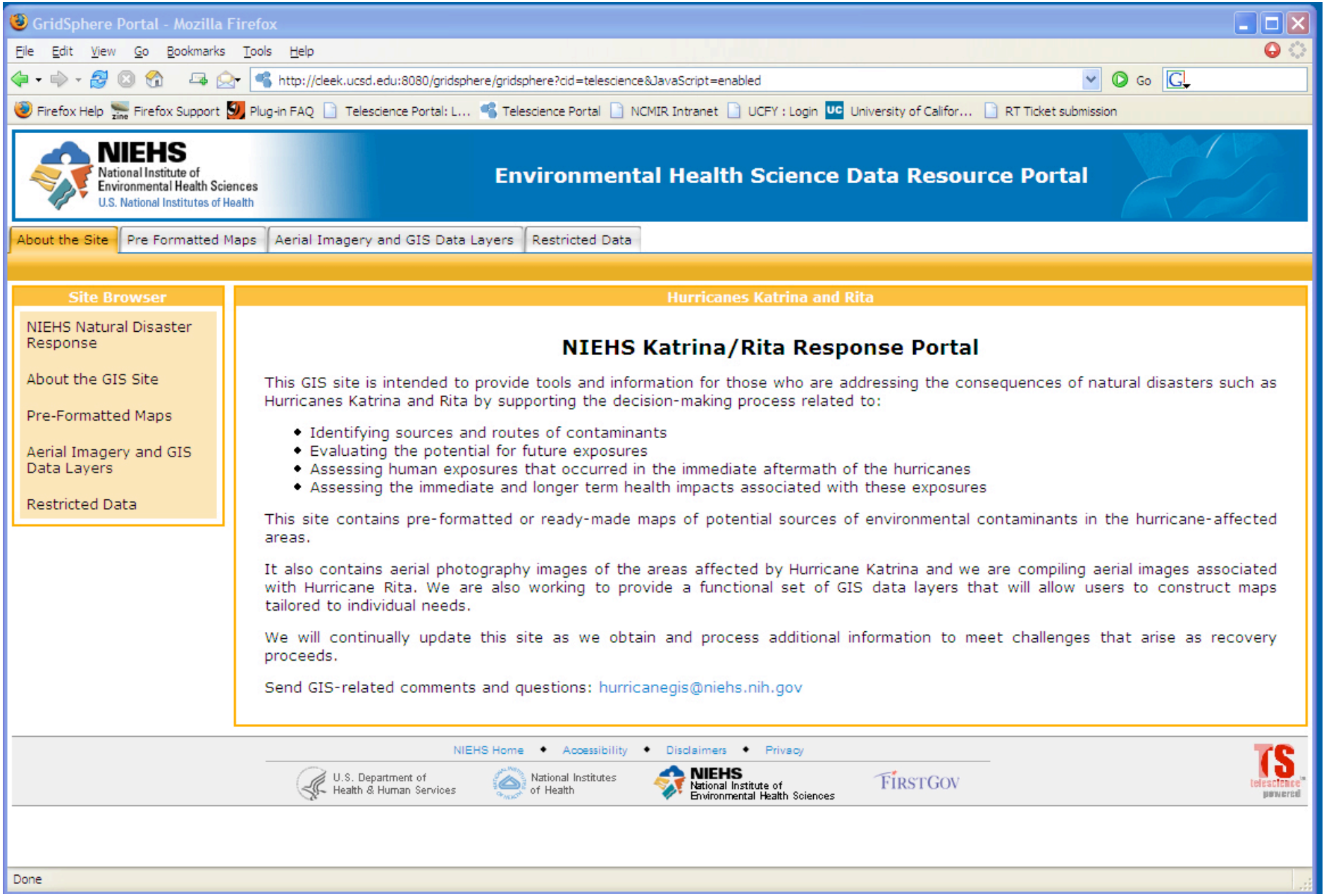
NPACI

UCSD

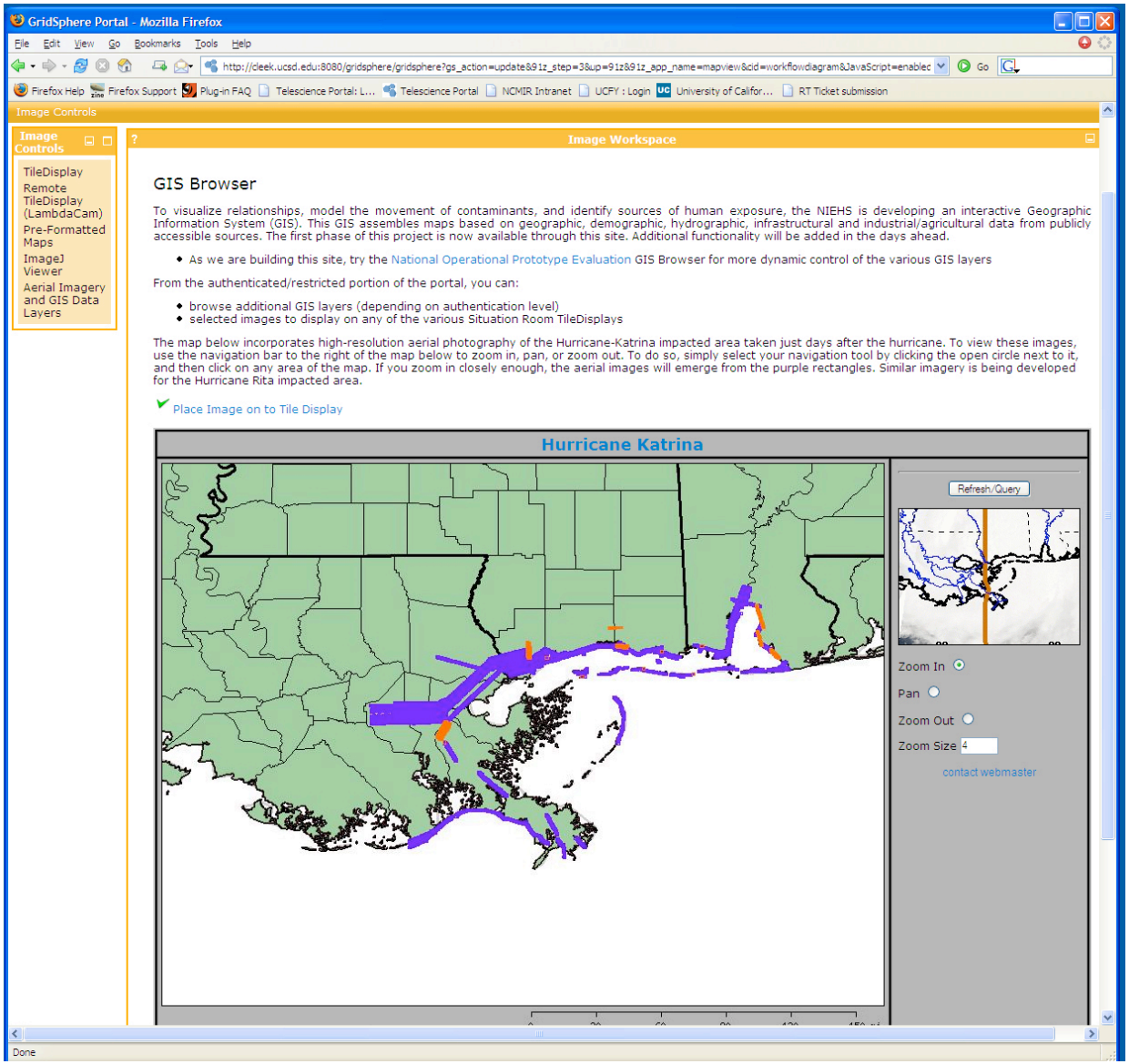
Telescience/ BIRN Portal Technology used as SARS Portal for Taiwan



Emergency Response: Created and functional in less than 24 hours

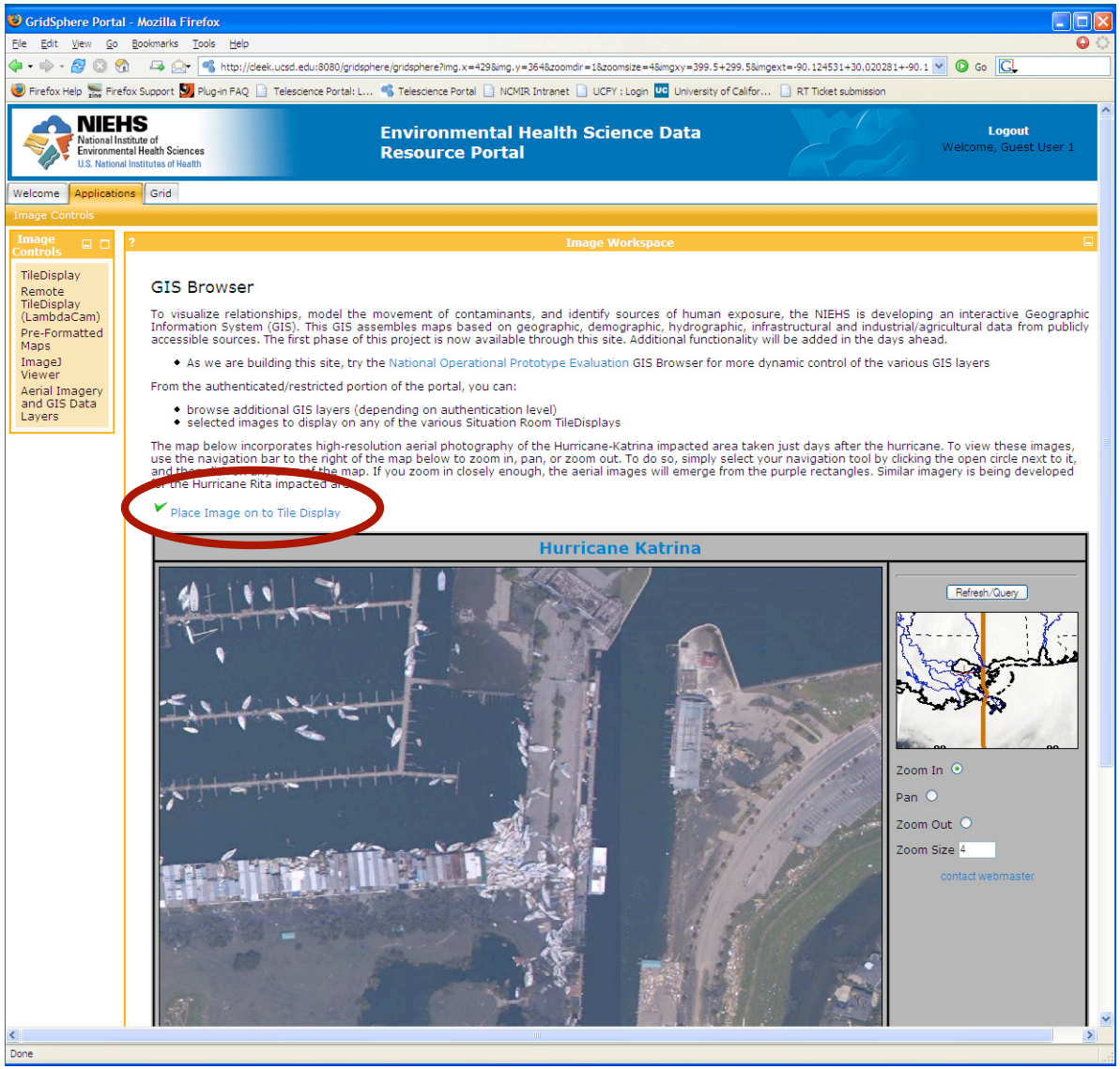

GridSphere Portal - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://cleek.ucsd.edu:8080/gridsphere/gridsphere?cid=telescience&JavaScript=enabled
Go
Firefox Help Firefox Support Plug-in FAQ Telescience Portal: L... Telescience Portal NCMIR Intranet UCFY : Login University of Califor... RT Ticket submission
NIEHS
National Institute of Environmental Health Sciences
U.S. National Institutes of Health
Environmental Health Science Data Resource Portal
About the Site
Pre Formatted Maps
Aerial Imagery and GIS Data Layers
Restricted Data
Site Browser
NIEHS Natural Disaster Response
About the GIS Site
Pre-Formatted Maps
Aerial Imagery and GIS Data Layers
Restricted Data
Hurricanes Katrina and Rita
NIEHS Katrina/Rita Response Portal
This GIS site is intended to provide tools and information for those who are addressing the consequences of natural disasters such as Hurricanes Katrina and Rita by supporting the decision-making process related to:
Identifying sources and routes of contaminants
Evaluating the potential for future exposures
Assessing human exposures that occurred in the immediate aftermath of the hurricanes
Assessing the immediate and longer term health impacts associated with these exposures
This site contains pre-formatted or ready-made maps of potential sources of environmental contaminants in the hurricane-affected areas.
It also contains aerial photography images of the areas affected by Hurricane Katrina and we are compiling aerial images associated with Hurricane Rita. We are also working to provide a functional set of GIS data layers that will allow users to construct maps tailored to individual needs.
We will continually update this site as we obtain and process additional information to meet challenges that arise as recovery proceeds.
Send GIS-related comments and questions: hurricanegis@niehs.nih.gov
NIEHS Home • Accessibility • Disclaimers • Privacy
U.S. Department of Health & Human Services
National Institutes of Health
NIEHS National Institute of Environmental Health Sciences
FIRSTGOV
Done

Integrated GIS Browser: View High Resolution Satellite Data, part 1



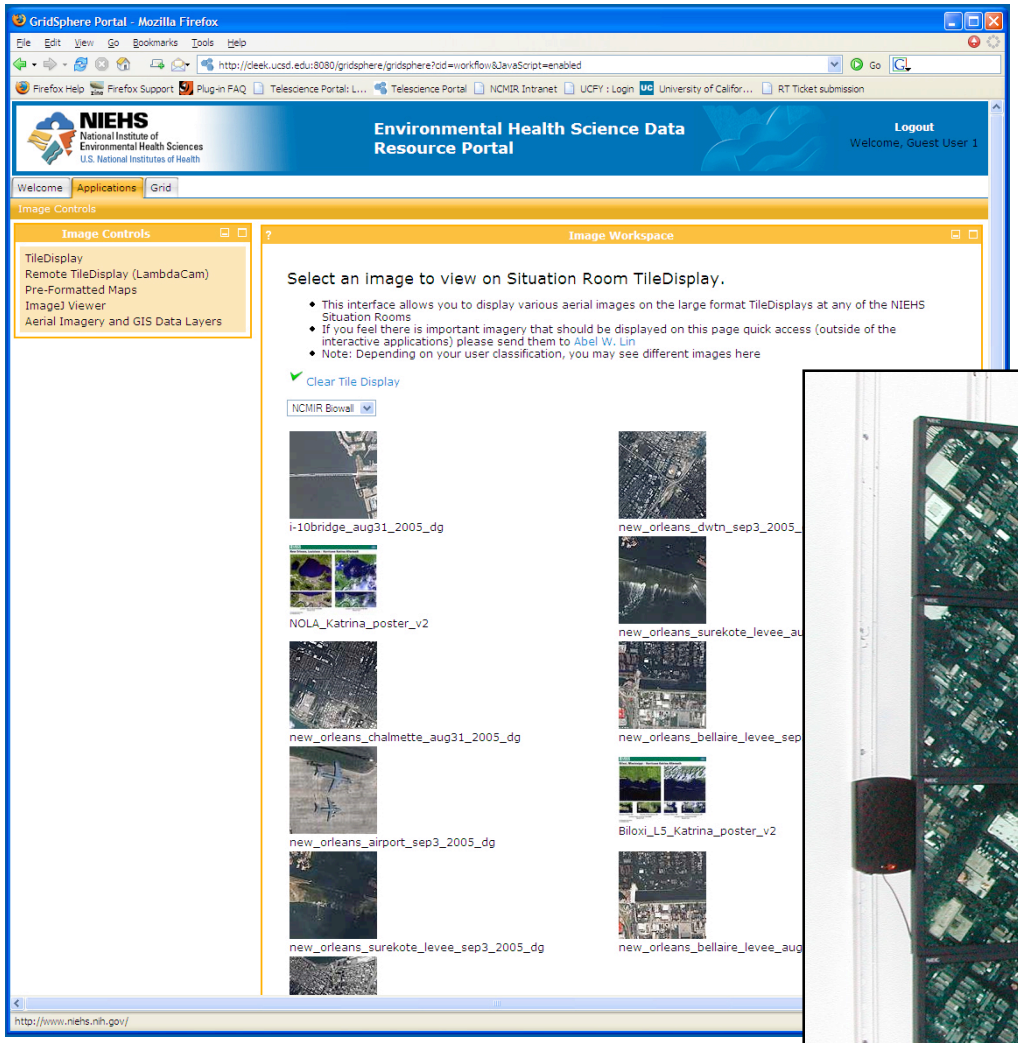
Integrated Browser to scan integrated satellite imagery

Integrated GIS Browser: View High Resolution Satellite Data, part 2

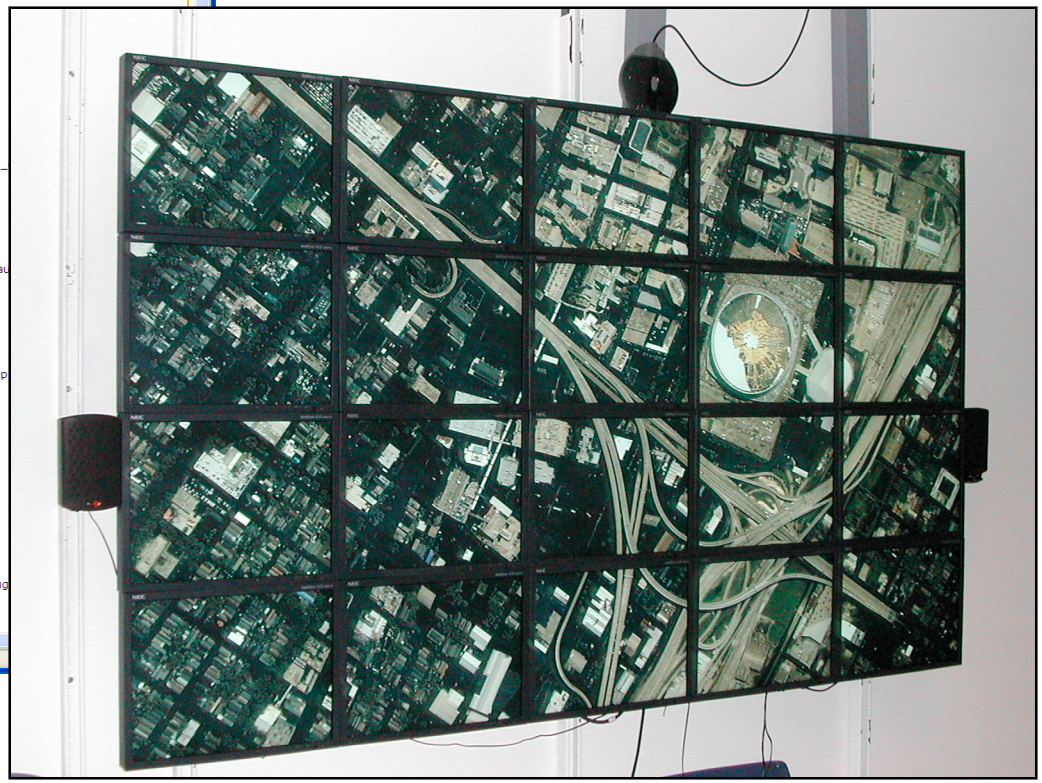


Can also Place GIS Image on TileDisplay for large format viewing

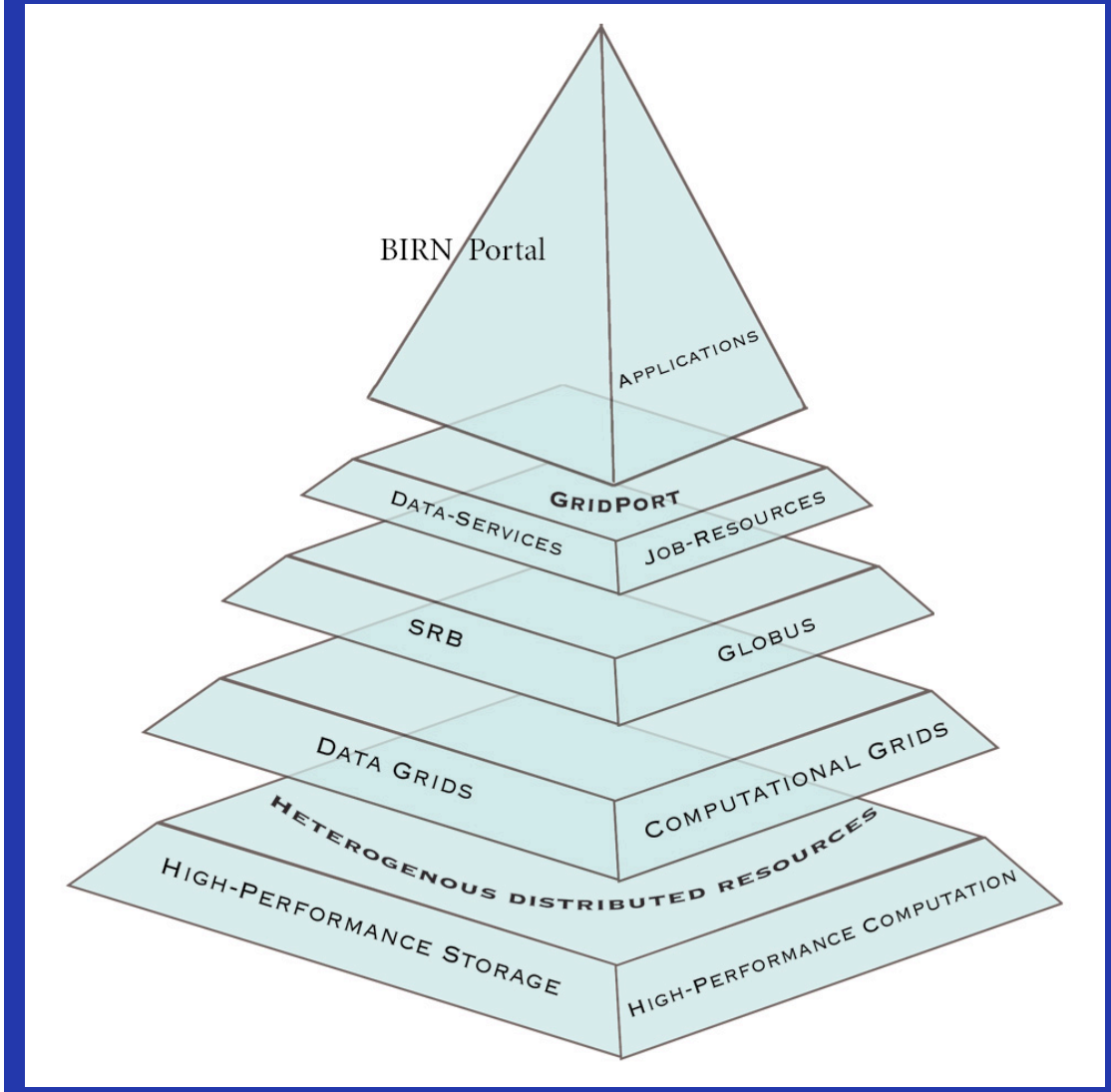
Launch images to large format TileDisplay visualization end-points



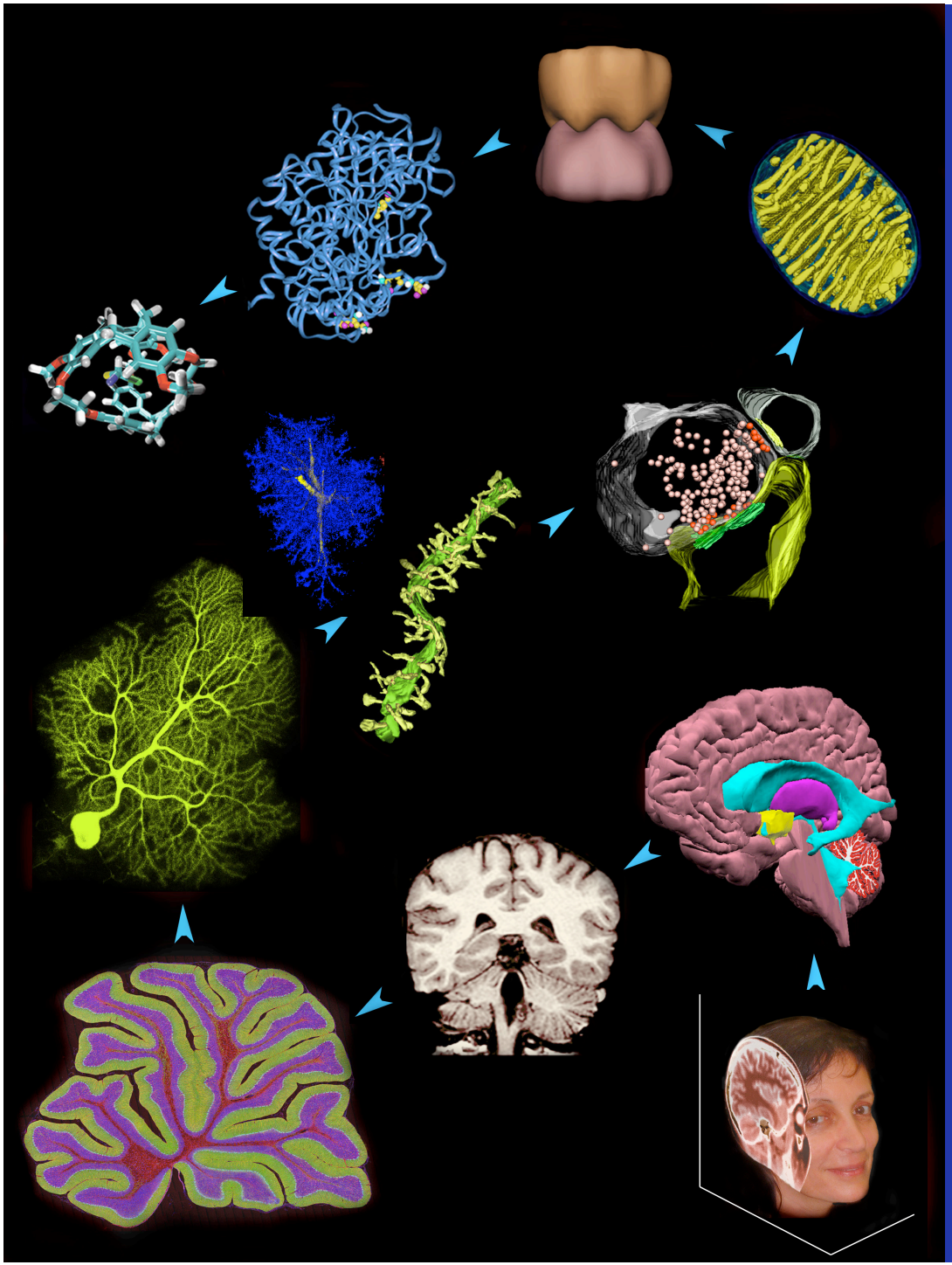
Put images on to various OptIPuter TileDisplay Visualization end-points by simply clicking thumbnails



Layered Architecture



- The Telescience Portal is composed of many "layers"
- Layers are modular, allowing for extension of any layer without great disruption to the entire system
- Every Layer has its own complexity and administration that was previously passed on to the end-user
- Portal centralizes all administrative details of each layer into a single username and pass phrase



Team Science Applied to Stretch Goals

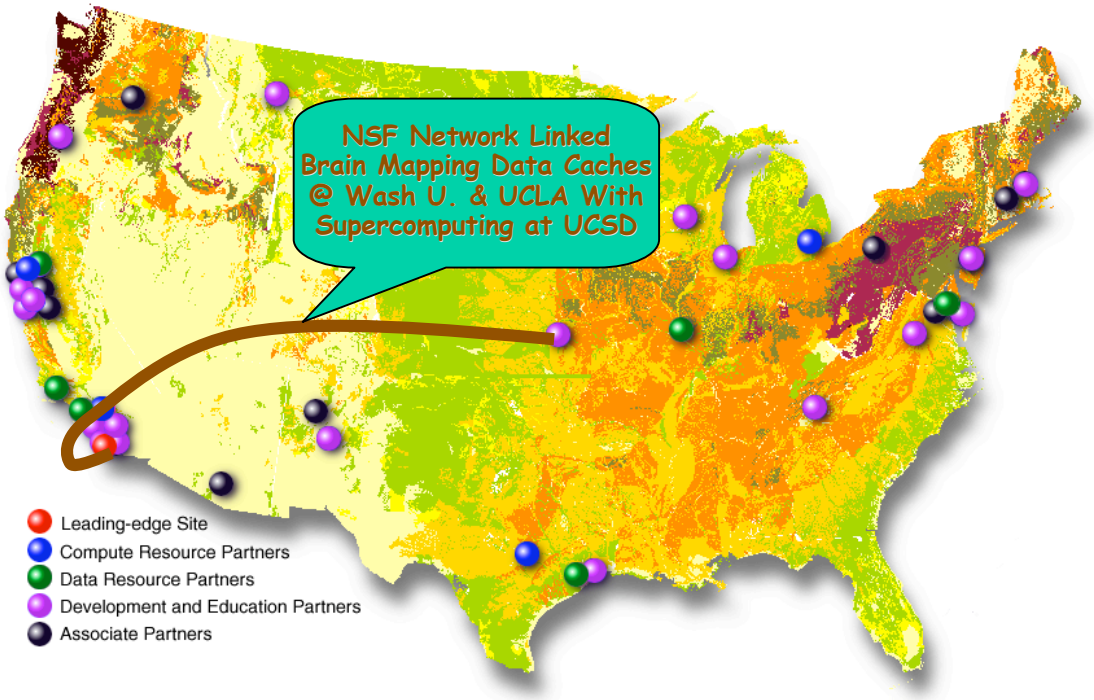
Enable new understanding of the brain by linking data about macroscopic brain function to its molecular and cellular underpinnings

- Federate Distributed Multiscale Brain Map Data
- Accommodate associated Large Scale Computational Challenges
- Provide Infrastructure for Construction of more Accurate Models and more Realistic Simulations of Brain Activity

Origins of IT Infrastructure used to build the BIRN:

Initiatives like the NSF - National Partnership for Advanced Computational Infrastructure (NPACI)

- ~50 partner sites
- Shared compute resources
- High-speed networks
- **Computational science efforts in ...**
  - *Neuroscience*
  - **Molecular Science**
  - **Earth Systems Science**
  - **Engineering**



- **Enabling Technologies:**
  - **Resources (TeraFlops, High Performance Networks, Data Caches)**
  - **Metacomputing (Grid Tools - Middleware)**
  - **Interaction Environments (Visualization - Science Portals)**
  - **Data-Intensive Computing (Databases - Data Integration)**

The NSF PACI Program Started in 1995
Current Program is "Cyberinfrastructure"

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

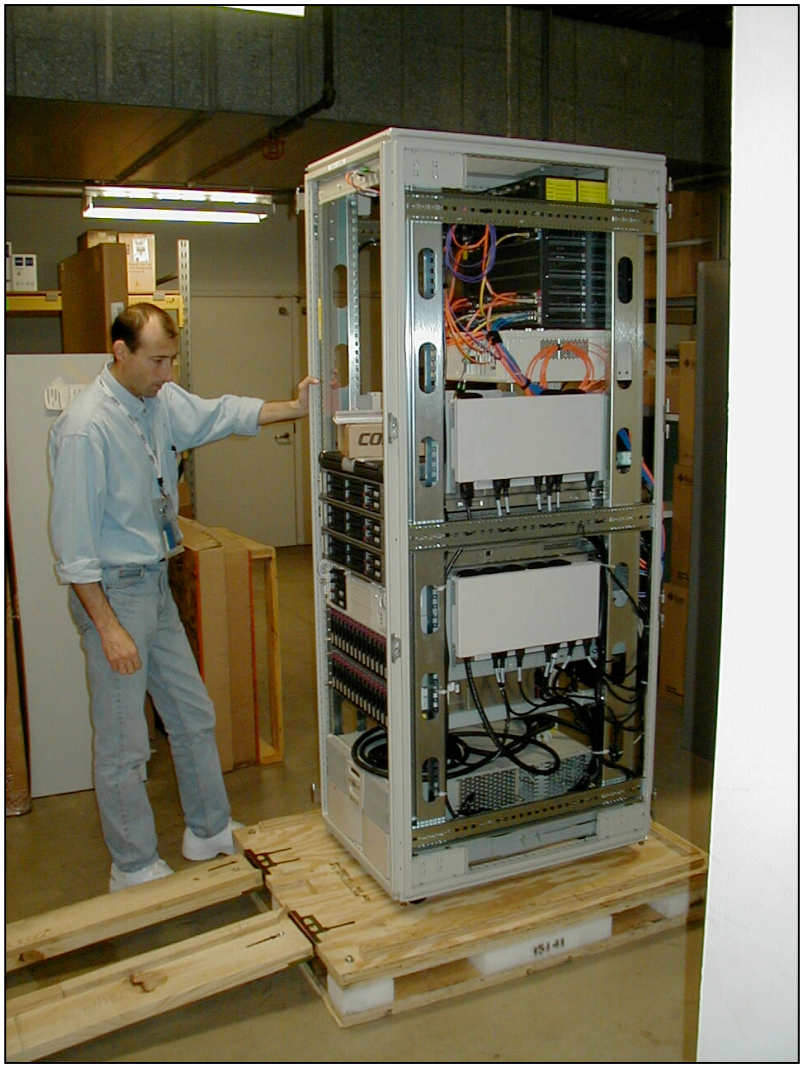
What is BIRN?

- Testbed for a biomedical knowledge infrastructure
- Creation and support federated bioscience databases
- Data integration
- Interoperable analysis tools
- Datamining software

✓ Scalable and extensible

✓ Driven by research needs pull, not technology push

We Began with Standard Hardware



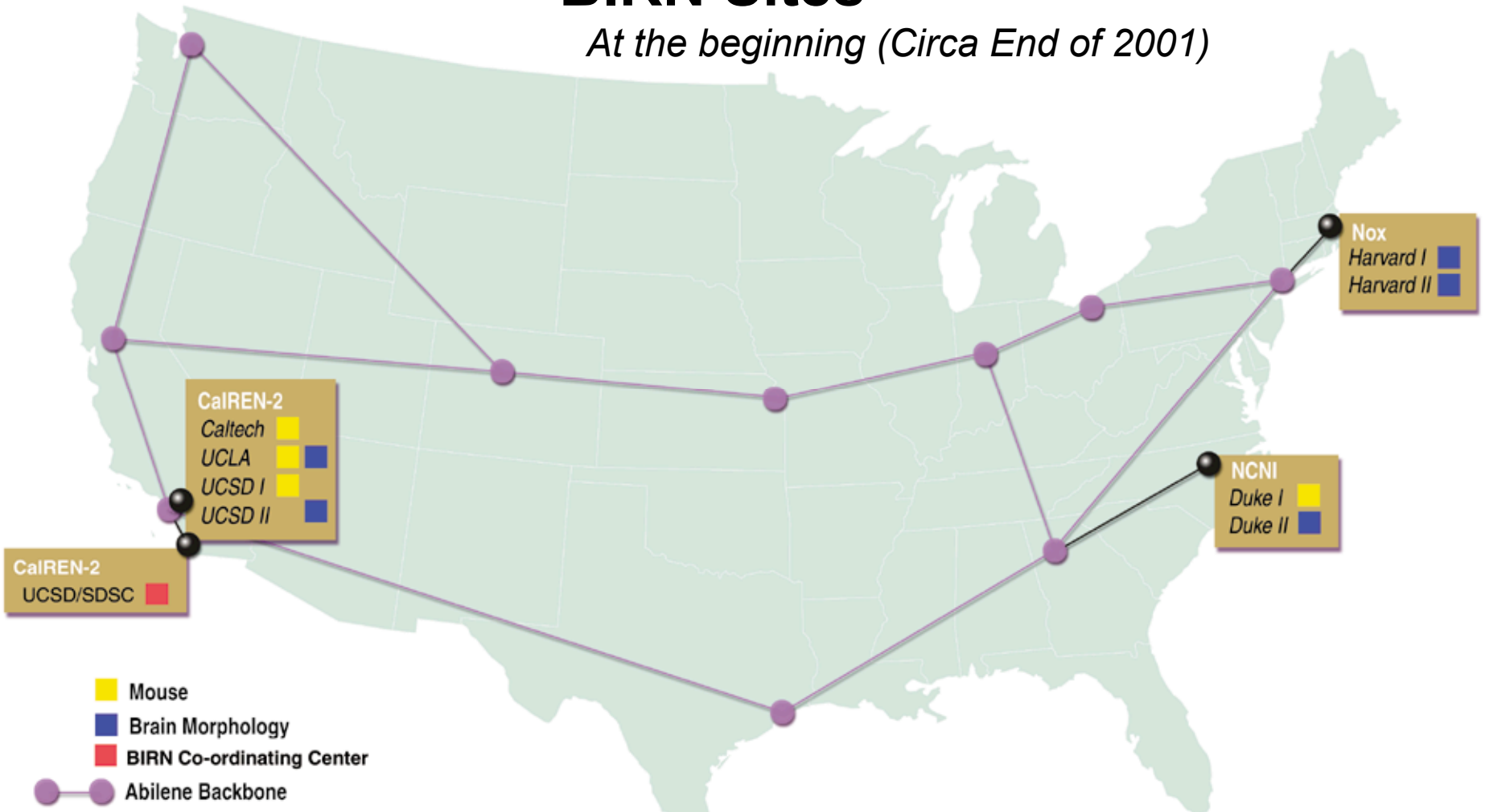
- This jumpstarted BIRN for functionality
- Software footprint is managed from the BIRN Coordinating Center
- Integration of domain tools, middleware, OS, updates, and more
- BIRN expansion/upgrade of existing sites has a more generic (and less expensive) hardware footprint

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

BIRN Must Accommodate Growth

BIRN Sites

At the beginning (Circa End of 2001)



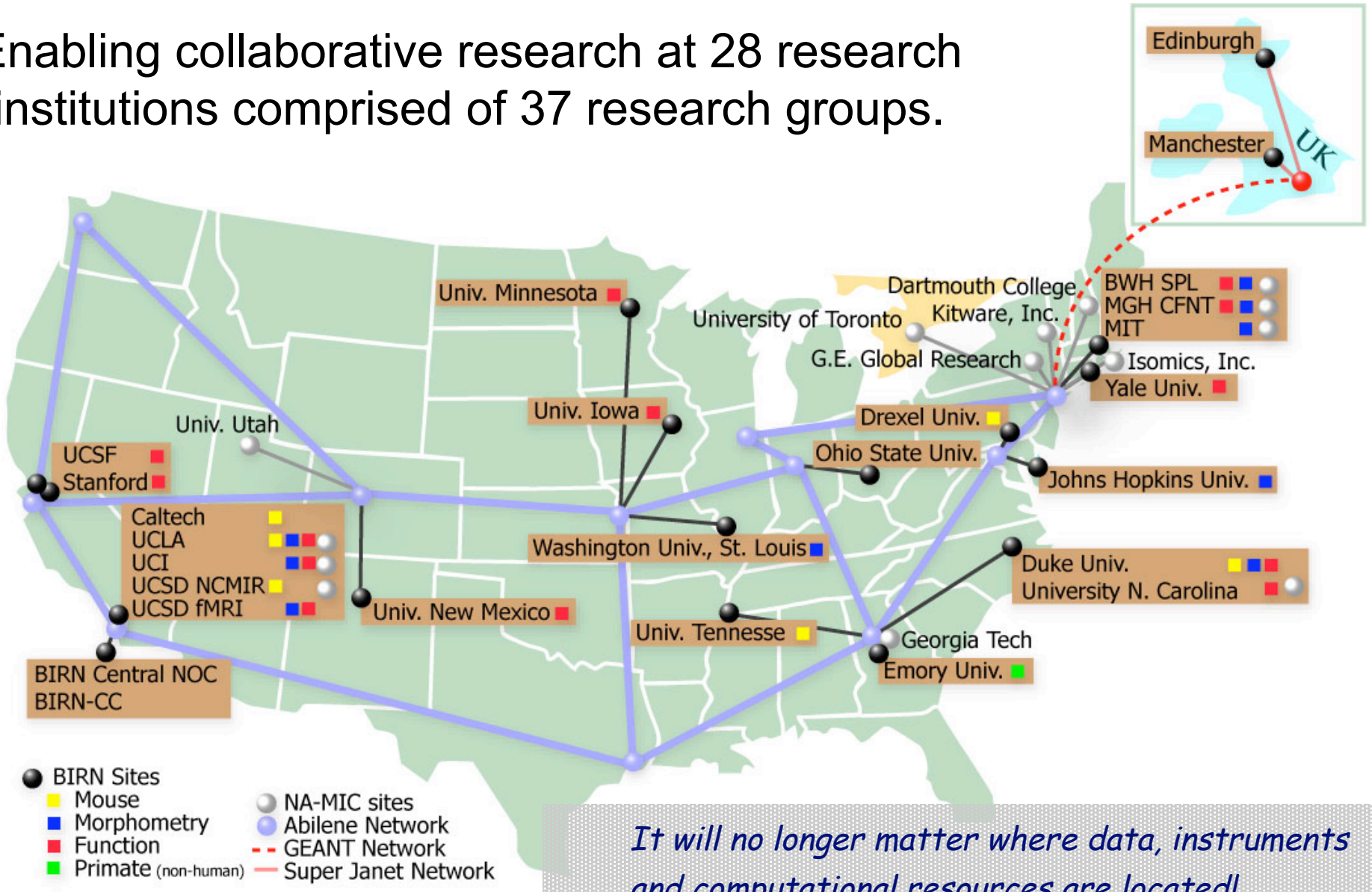
10+ Distinct Installations, ~ 100 Individual Machines

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

From the Expanding the BIRN Meeting @ NCRR: December. 6 & 7, 2001)

The BIRN Collaboratory Today

Enabling collaborative research at 28 research institutions comprised of 37 research groups.

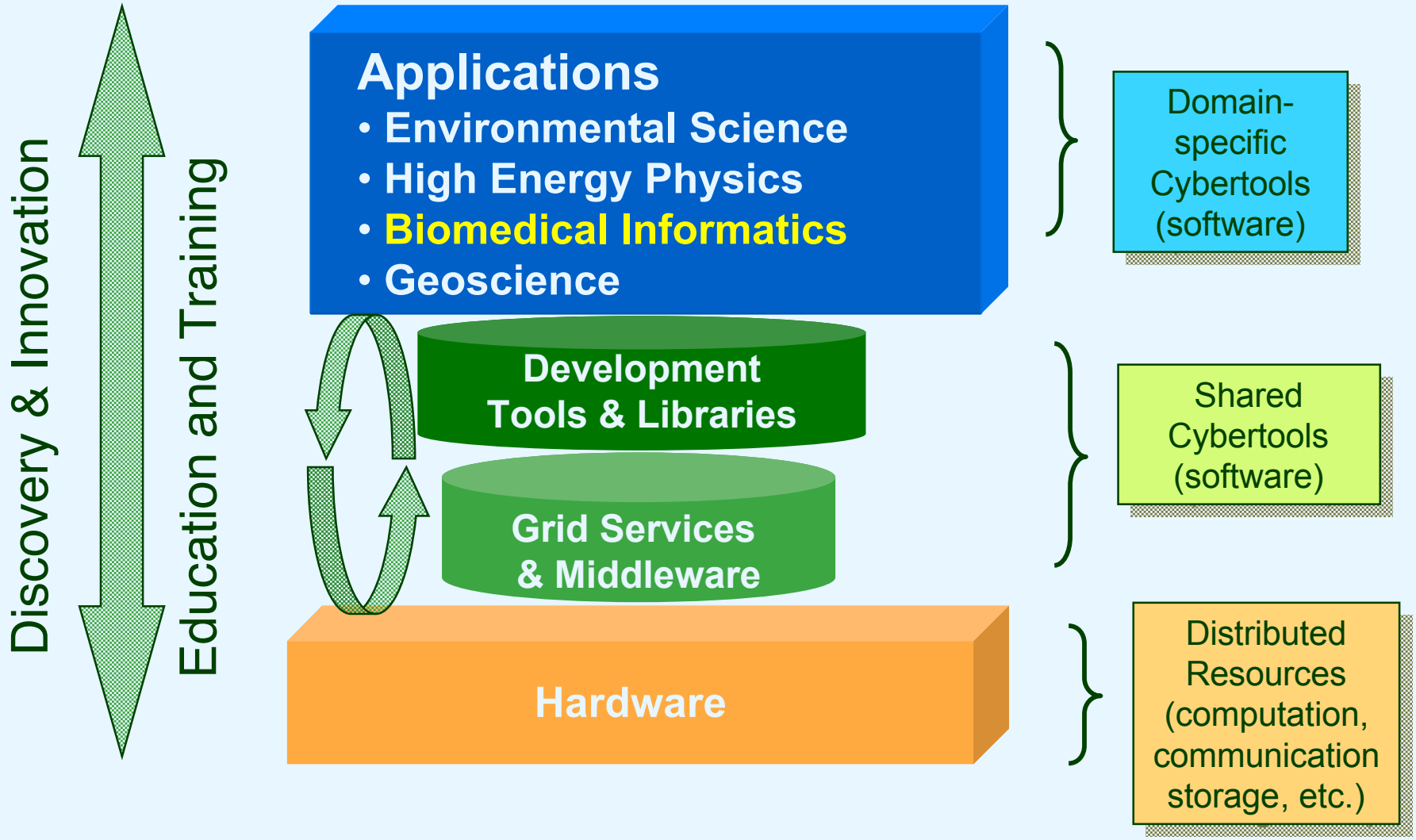


It will no longer matter where data, instruments and computational resources are located!

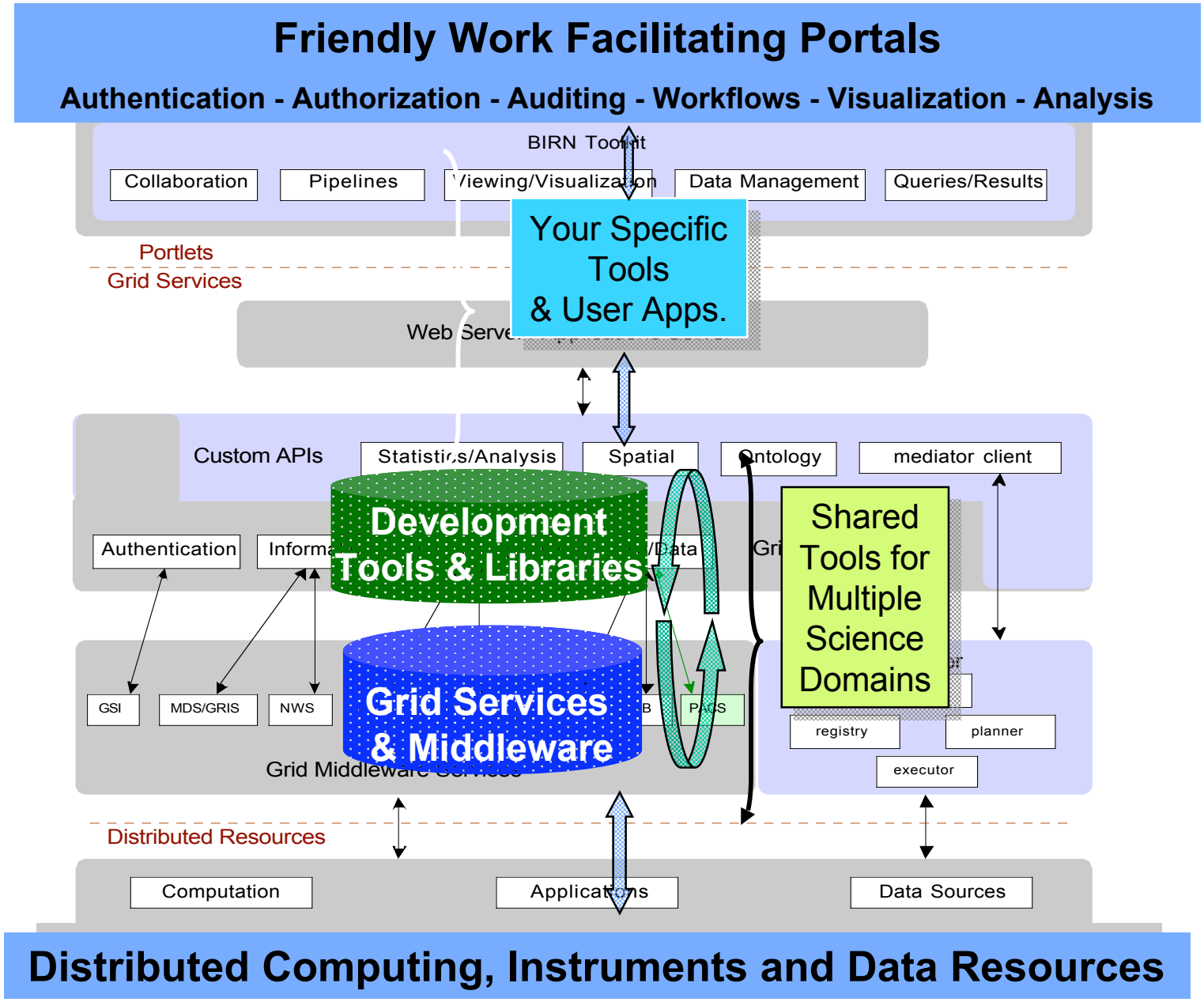
Integrated Cyberinfrastructure System

meeting the needs of multiple communities

Source: Dr. Deborah Crawford, Chair, NSF ***CyberInfrastructure*** *Working Group*



BIRN Core Software Infrastructure



- BIRN builds on evolving community standards for middleware
- Adds new capabilities required by projects
- Does System Integration of domain-specific tools building a distributed infrastructure
- Utilizes commodity hardware and stable networks for baseline connectivity

Software Problem in a Nutshell

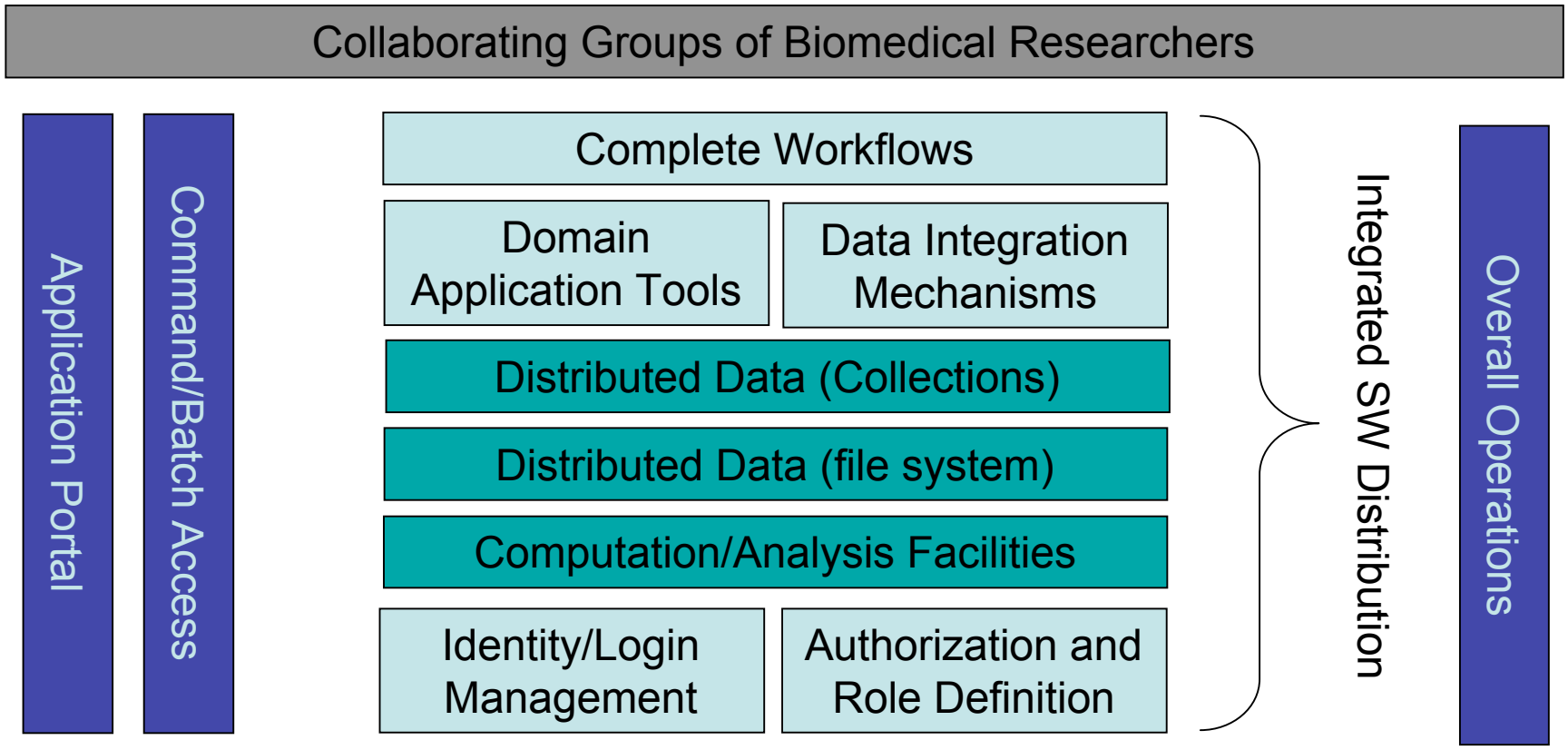
- **Enable Analysis of Distributed Biomedical Data in a National-Scale Production Facility**

Data & Network	• Data sets are large – Data sets are many • Enable new queries that integrate multiple sources • Specialized application codes (from Test Beds) need to work on BIRN-accessible data
CPU	• Some analysis pipelines require significant computation
Security	• Privacy, patient anonymity required • Institutional ownership of originals

- **Easily Replicate Entire Software Stack** (Including Centralized Services) for other Groups

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

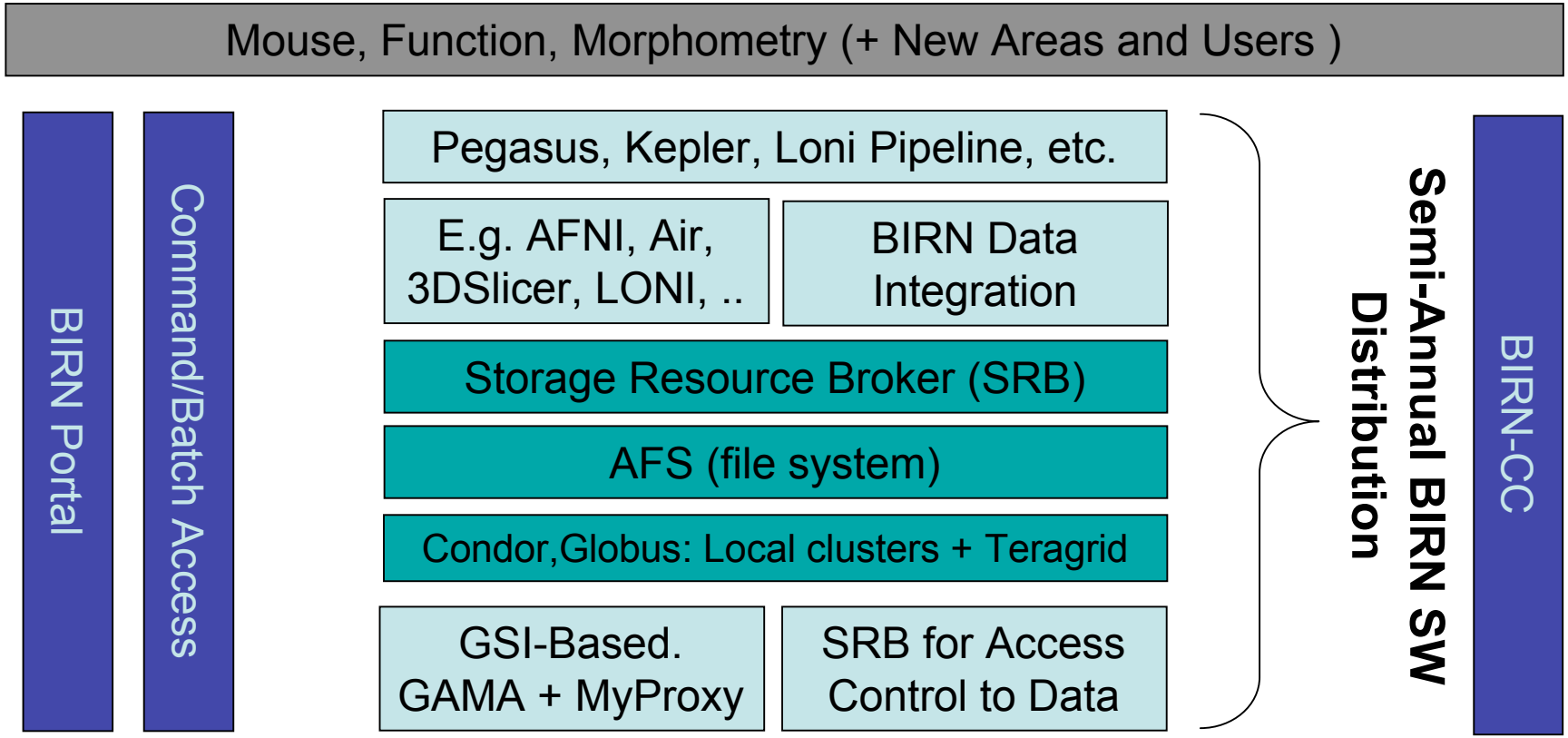
Major System Components



Note: Similar Structure to Many Other Grids

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

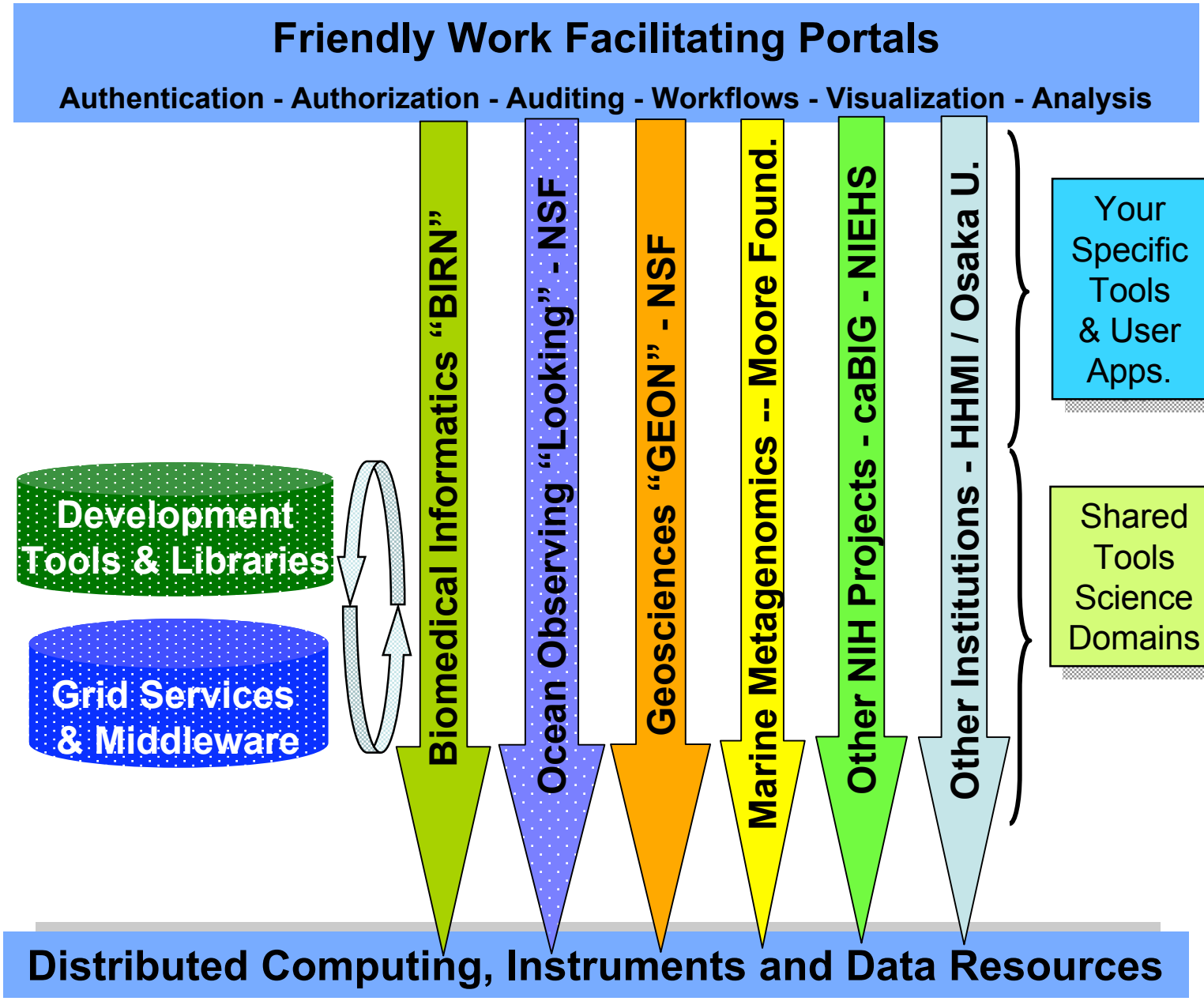
Specific Implementations



Note: Similar Structure as Many Other Grids

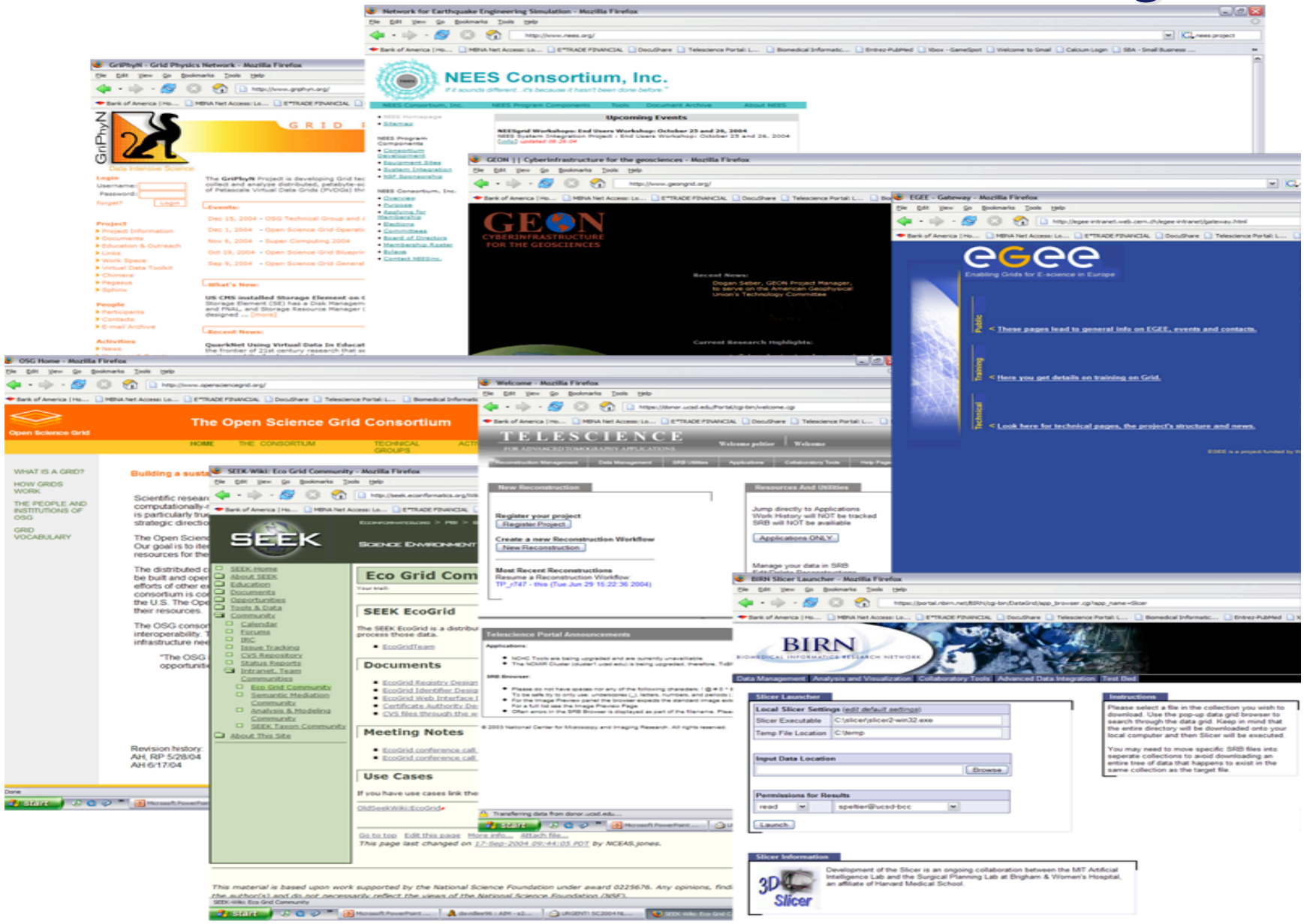
BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

BIRN Core Software Infrastructure



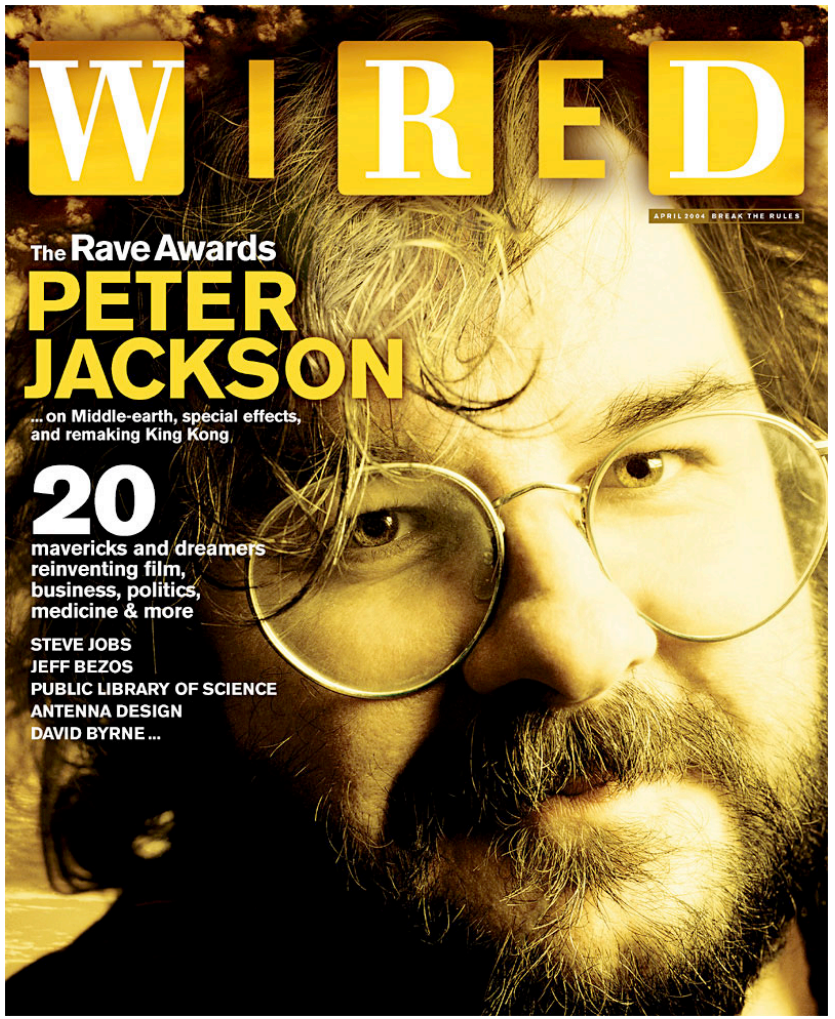
- BIRN CC builds on evolving standards for portals and middleware
- Adds new capabilities required by projects
- Provides system integration of domain-specific tools building a distributed infrastructure
- Utilizes commodity hardware and stable networks for baseline connectivity

The Grid is becoming the backbone for collaborative science and data sharing



Grid Infrastructure in Action

- **The Grid is already having an impact…**
  - **Many projects in many subjects:**
    - Life sciences
    - Medicine
    - Environment
    - Engineering
    - Materials
    - Chemistry
    - Physics
  - **BIRN embodies the most innovative use of data, metadata & portals**



BIRN cited as successful model of grid computing.

BIRN has the Advantage of having Developed an "End-to-End" Infrastructure: ***Built around research projects with geographically distributed data.***

- **Consists of all the components required to effectively share and collaboratively explore data**
  - **The BIRN Rack (BIRN site infrastructure)**
  - **The BIRN Portal**
  - **The BIRN Virtual Data Grid**
  - **The BIRN Data Integration Infrastructure**
  - **The BIRN Computational GRID**
- **The system integration, development, deployment and management of this infrastructure is the main focus of activities within the BIRN Coordinating Center**

The BIRN Portal

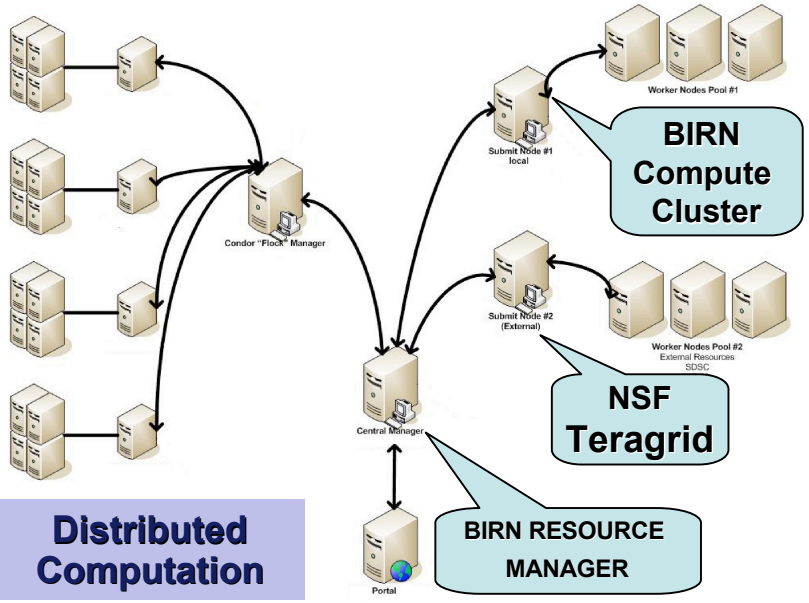
- Application environment that provides transparent and pervasive access to the BIRN infrastructure (i.e. tools, applications, resources) with a **Single Login** from any Internet capable location

- Provides simple, intuitive access to distributed resources for data storage, distributed computation, and visualization

- Provides a scalable interface for users of all backgrounds and levels of expertise

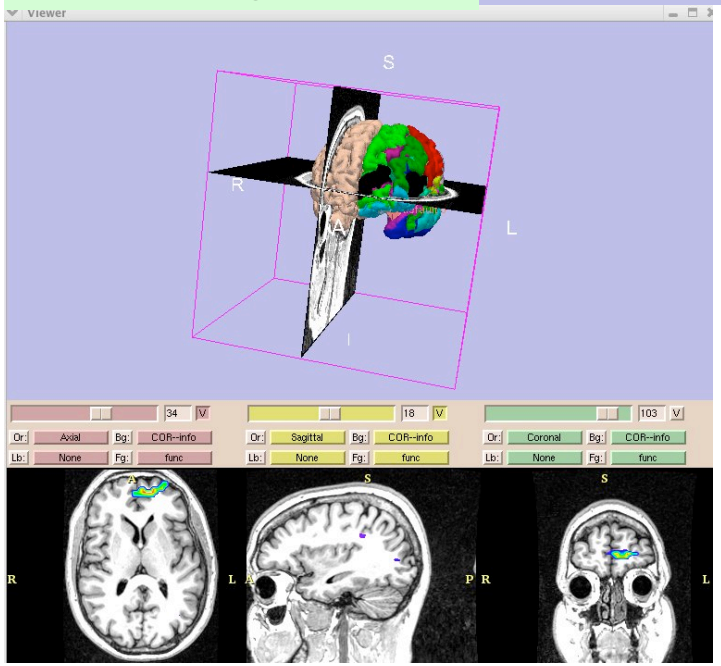
The BIRN Portal Provides an Intuitive Interface to Software Tools, Data and Computational Resources in the BIRN Collaboratory

The BIRN Portal
Launches from any Internet Connected Desktop, Laptop, PDA



BIRN VIEWERs - eg., SLICER; ImageJ, MBAT

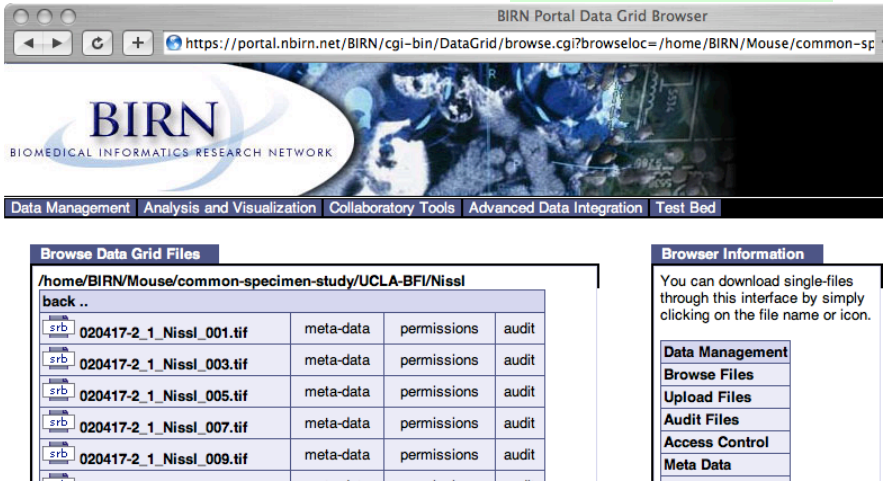
Data Visualization



Distributed Computation

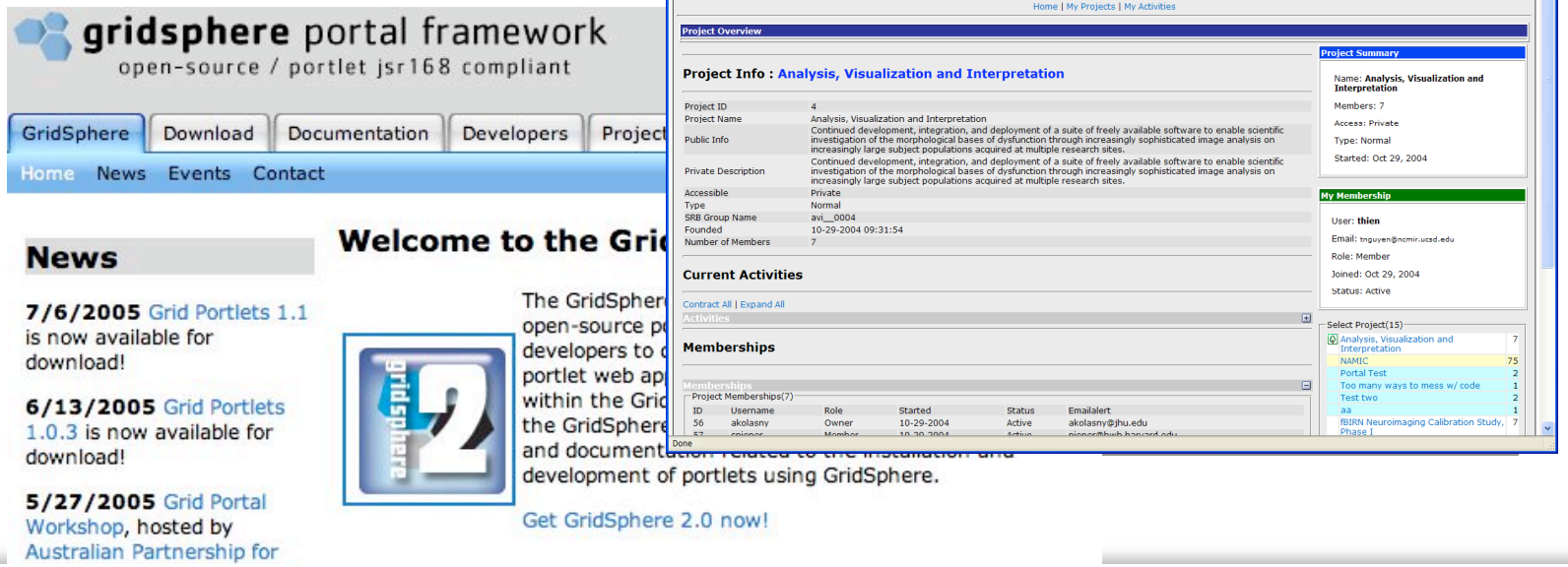
Data Management

Browse Data Grid Files



BIRN is a Leader in Portal Technology

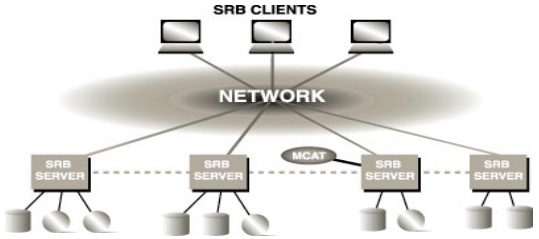
- **The BIRN-CC is supporting development of the leading open-source standards-based grid portal**
- **Application environment that provides transparent and pervasive access to the BIRN infrastructure (i.e. tools, applications, resources) with a Single Login from any Internet capable location**
- **Support for dynamic collaborative projects**



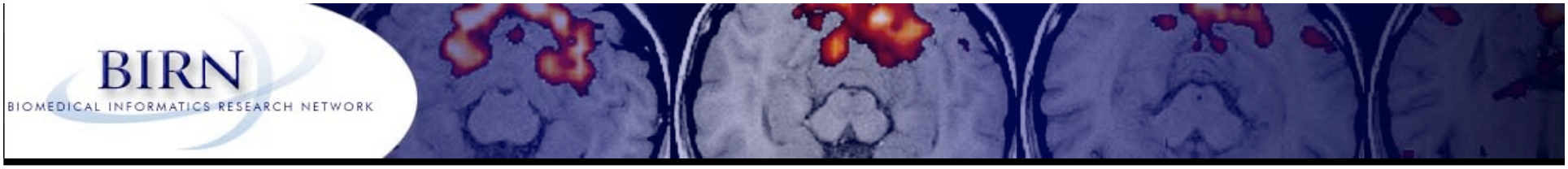
BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

Benefits of a Data Grid

- **Uniform interface for connecting to heterogeneous distributed data resources**
  - *Allows for any "grid enabled" tool to interact with data no matter where it is located or what it is located on*
- **Allows for the seamless creation and management of distributed data sets**
  - *Distributed data appear as a single managed collection both to users and tools*
- **Access is Managed using GRID Authentication through BIRN Portal**



BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK



- BIRN Coordinating Center (UCSD)
- BIRN Test Bed Projects
  - Morphometry BIRN
  - Function BIRN
  - Mouse BIRN

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

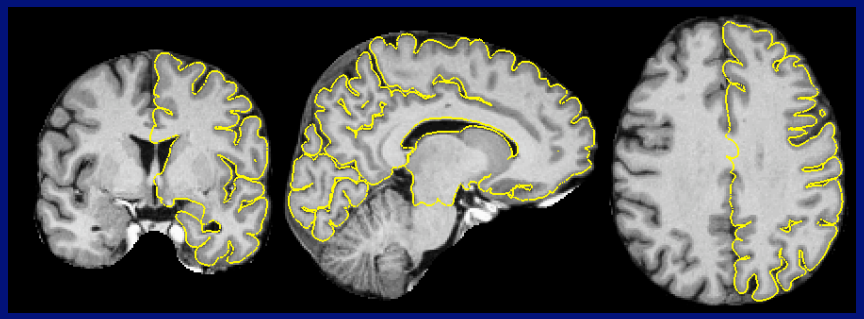
Brain Morphometry BIRN

Clinical Aims

- Do structural differences contribute to specific symptoms such as memory dysfunction or depression?
- Do specific structural differences distinguish specific diagnostic categories?

Technological Aims

- Attempt to overcome obstacles to the use of neuroimaging data as quantitative outcome measures for clinical investigation including the issues raised by longitudinal and multi-site studies.



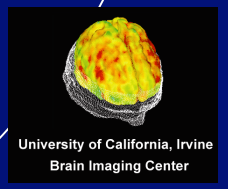
BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

Morph BIRN

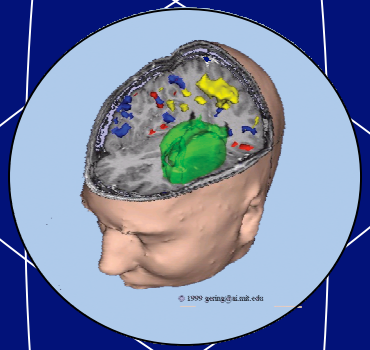
UCLA

MASSACHUSETTS GENERAL HOSPITAL

UCSD-GCRC
FMRI Center



University of California, Irvine
Brigham Imaging Center



SPL

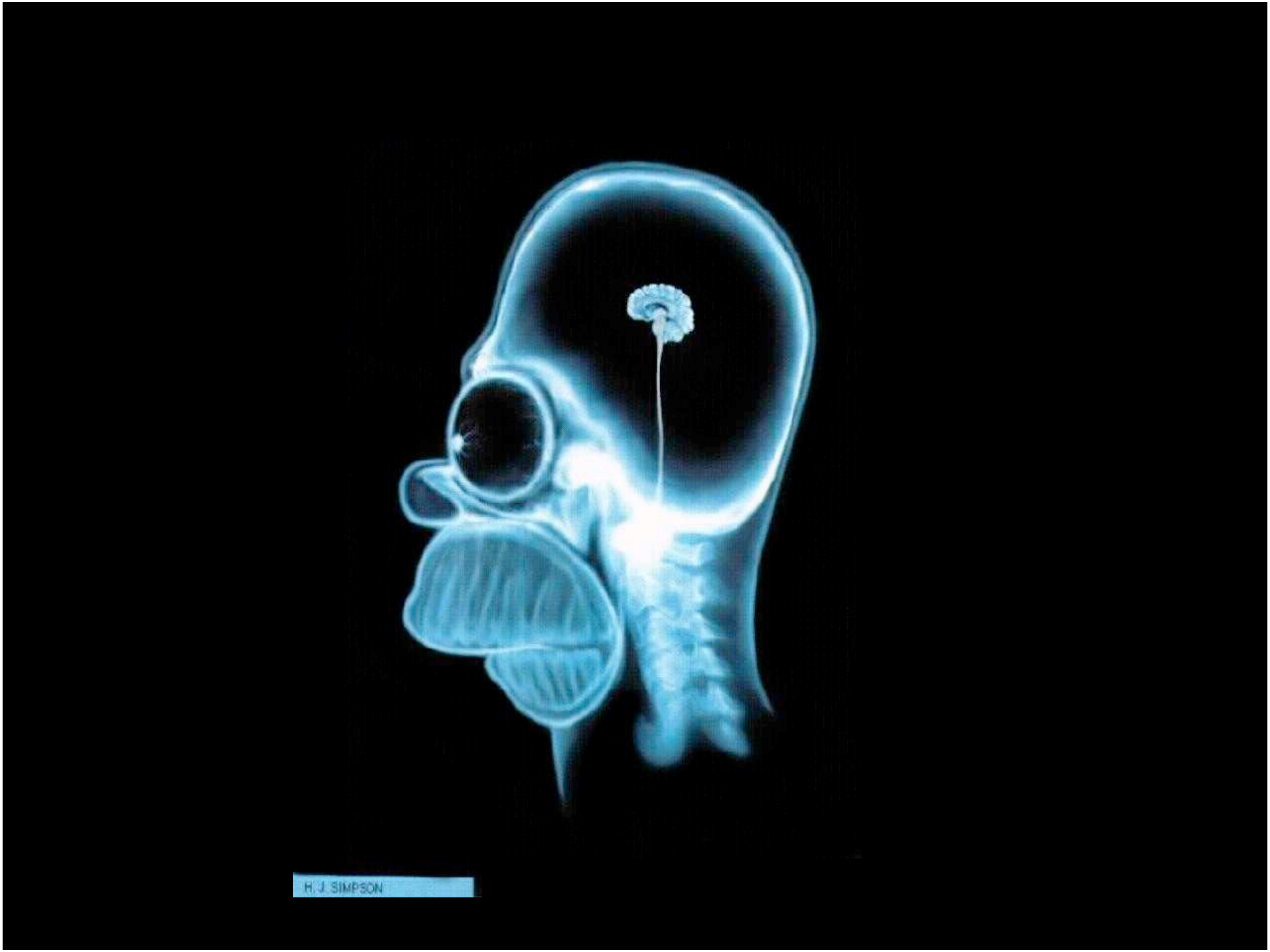
Brigham
and Women's
Hospital

Duke University
Medical School

UNC
SCHOOL OF MEDICINE

Center for
IMAGING
SCIENCE
JOHNS HOPKINS UNIVERSITY

Johns Hopkins
University

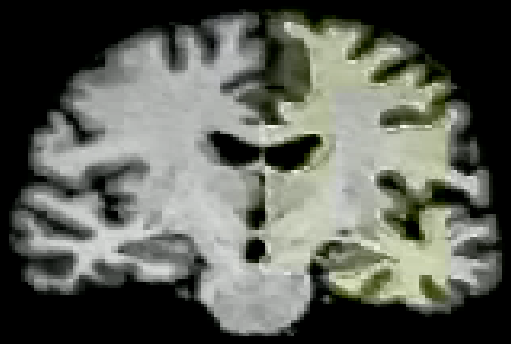

H. J. SIMPSON

Human Subjects Considerations

- High-resolution structural images can be used as an identifier.
  - Reconstruction of face from raw anatomical data might be able to be used to identify subject
  - Some members of BIRN require/desire unaltered raw data
  - BIRN has received approval from local IRBs to allow for the sharing of raw anatomical data from prospective subjects with current authorized BIRN members (i.e. fBIRN Human Phantoms)
- BIRN must conform to multiple overlapping regulations
  - Common Rule
  - HIPAA
  - State Law

Cortical Thickness Estimation with Sub-Voxel Accuracy

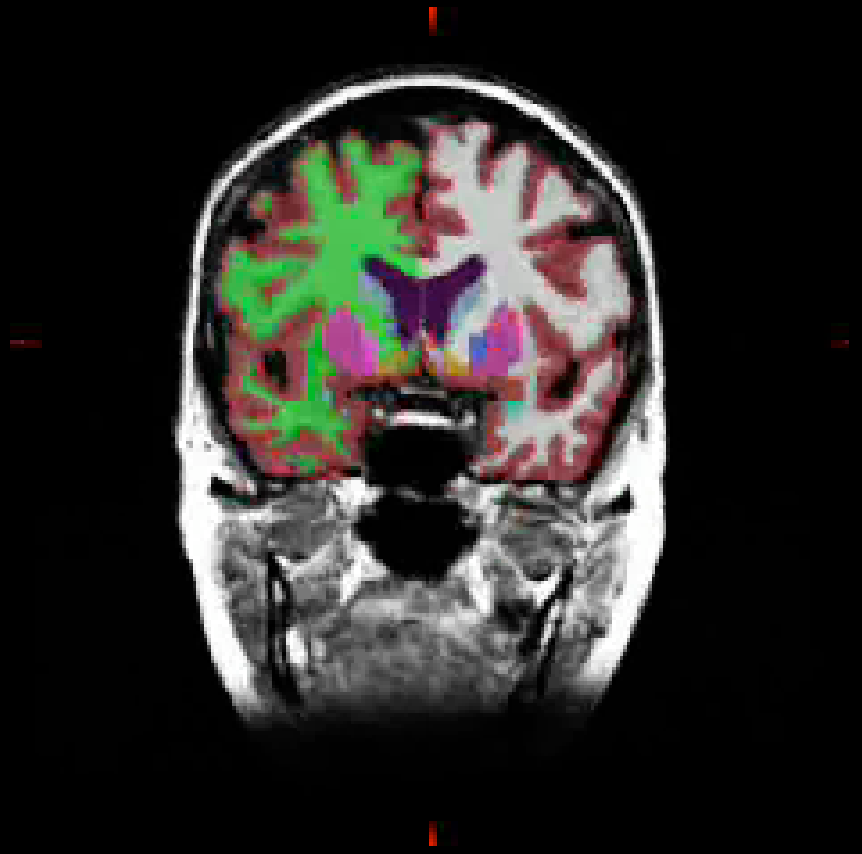
Gray-white boundary



Pial surface

From Anders Dale / Mass General Hospital - Harvard

Automated Whole-Brain Segmentation

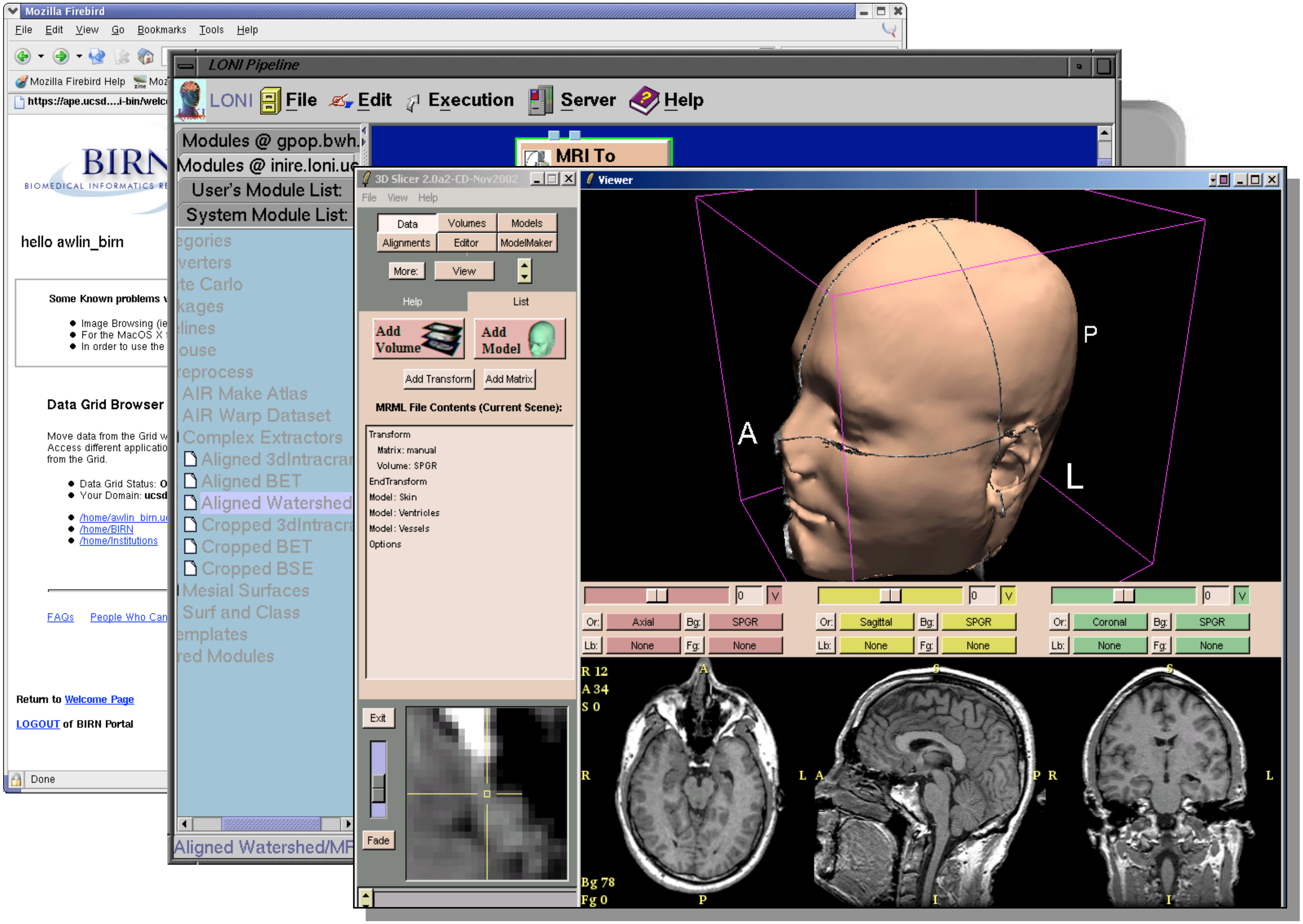


Part of Free Surfer

By Bruce Fishl
and
Anders Dale
(MGH)

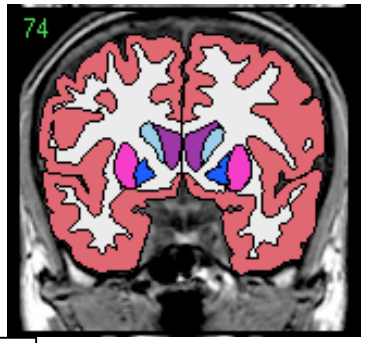
- Cerebellar cortex
- Cerebellar WM
- 4th ventricle
- RH cerebral WM
- LH cerebral WM
- Hippocampus
- LH pallidum
- Thalamus
- Cerebral cortex
- Misc.
- Lateral ventricle
- Caudate

BIRN Portal: Launches Scientific Workflow



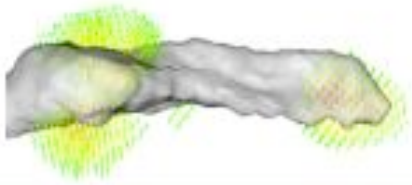
[illegible] LONI settings

SASHA Project



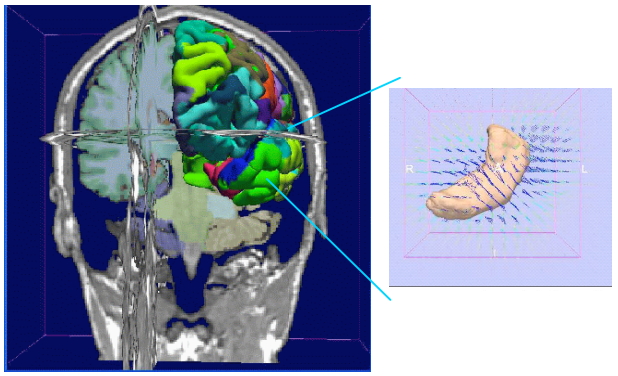
3

MGH
Segmentation



4

JHU
Shape Analysis
of Segmented Structures



5

BWH
Visualization

TeraGRID

1

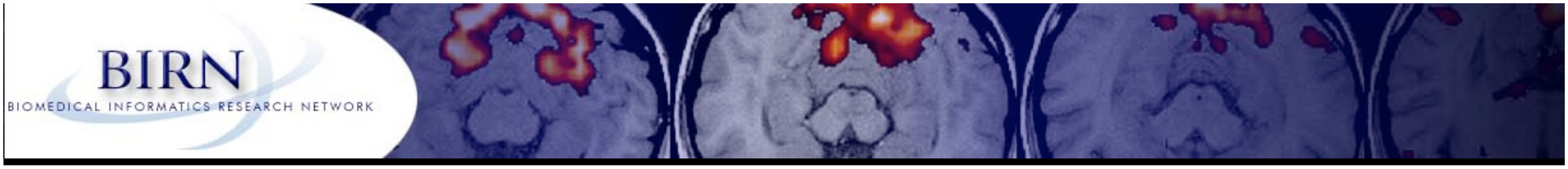
Data Donor Sites

SRB

De-identification
And upload

2

Goal: comparison and quantification of structures' shape and volumetric differences across patient populations



- A team of eleven universities studying regional brain dysfunctions related to the progression and treatment of ***schizophrenia.***
  - Prefrontal Cortex (executive dysfunction)
  - Temporal Lobe (auditory processing)

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

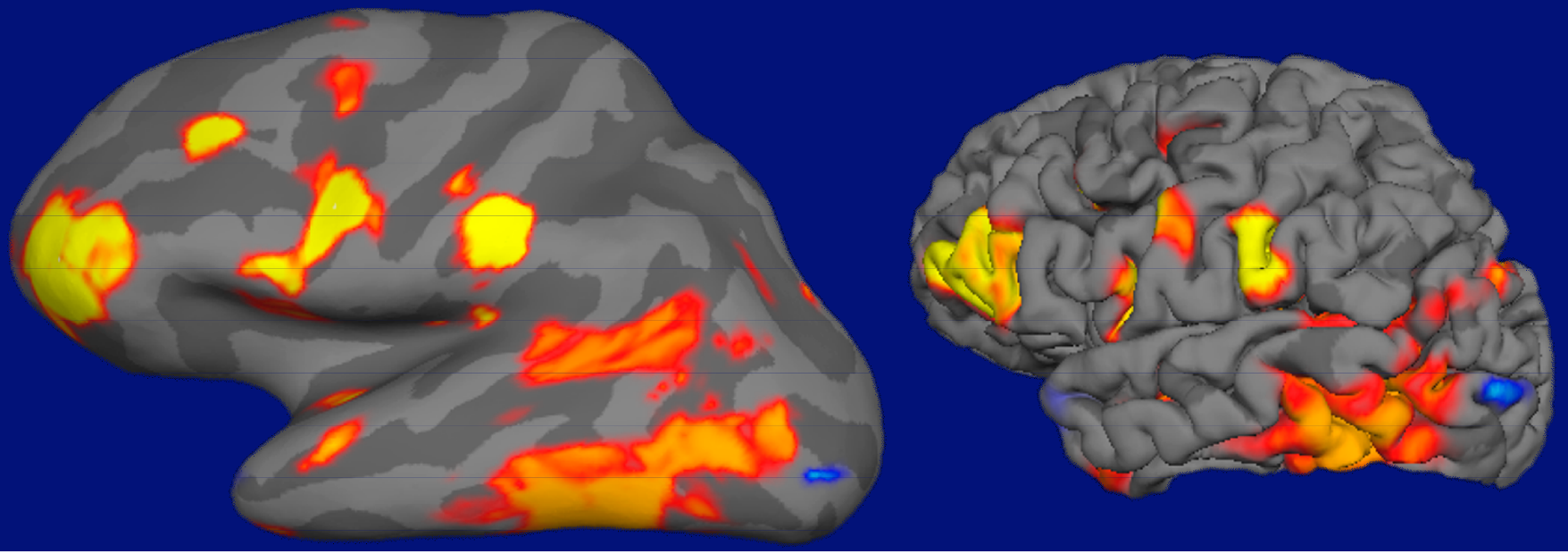
Functional Imaging

Clinical Aims

- Is Frontal and Temporal Lobe Dysfunction the Cause of Schizophrenia?
- How can Treatment Reverse this Dysfunction?

Technological Aims

- Integration of 4D Data from Multiple Sites - Acquired with Different Non-Invasive Imaging Devices
- Integration of Information Obtained with Different Brain Activation Tasks.



Function BIRN

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

Stanford

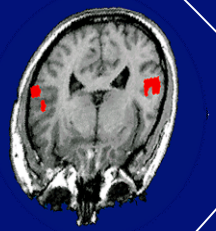
MASSACHUSETTS GENERAL HOSPITAL

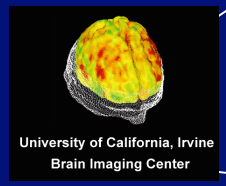
UCSD-GCRC FMRI Center

UNC SCHOOL OF MEDICINE

Brain Imaging and Analysis Center

The University of New Mexico Health Sciences Center





Duke University Medical School

nirl

CMRR
University of Minnesota

UNIVERSITY of IOWA HEALTH CARE
MENTAL HEALTH CLINICAL RESEARCH CENTER

LONI

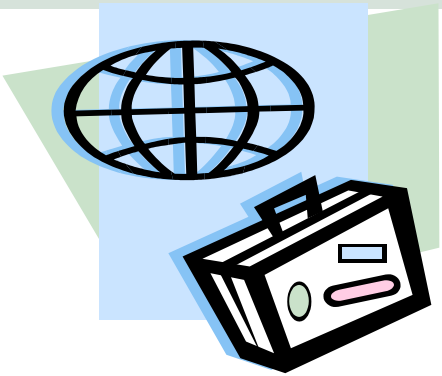
UCLA

SPL

Brigham and Women's Hospital

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

Traveling Humans Study



- Subjects traveled around the country to be scanned at all Function BIRN sites
- Unique dataset: (Subject) x (site) interactions can be measured for the first time

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

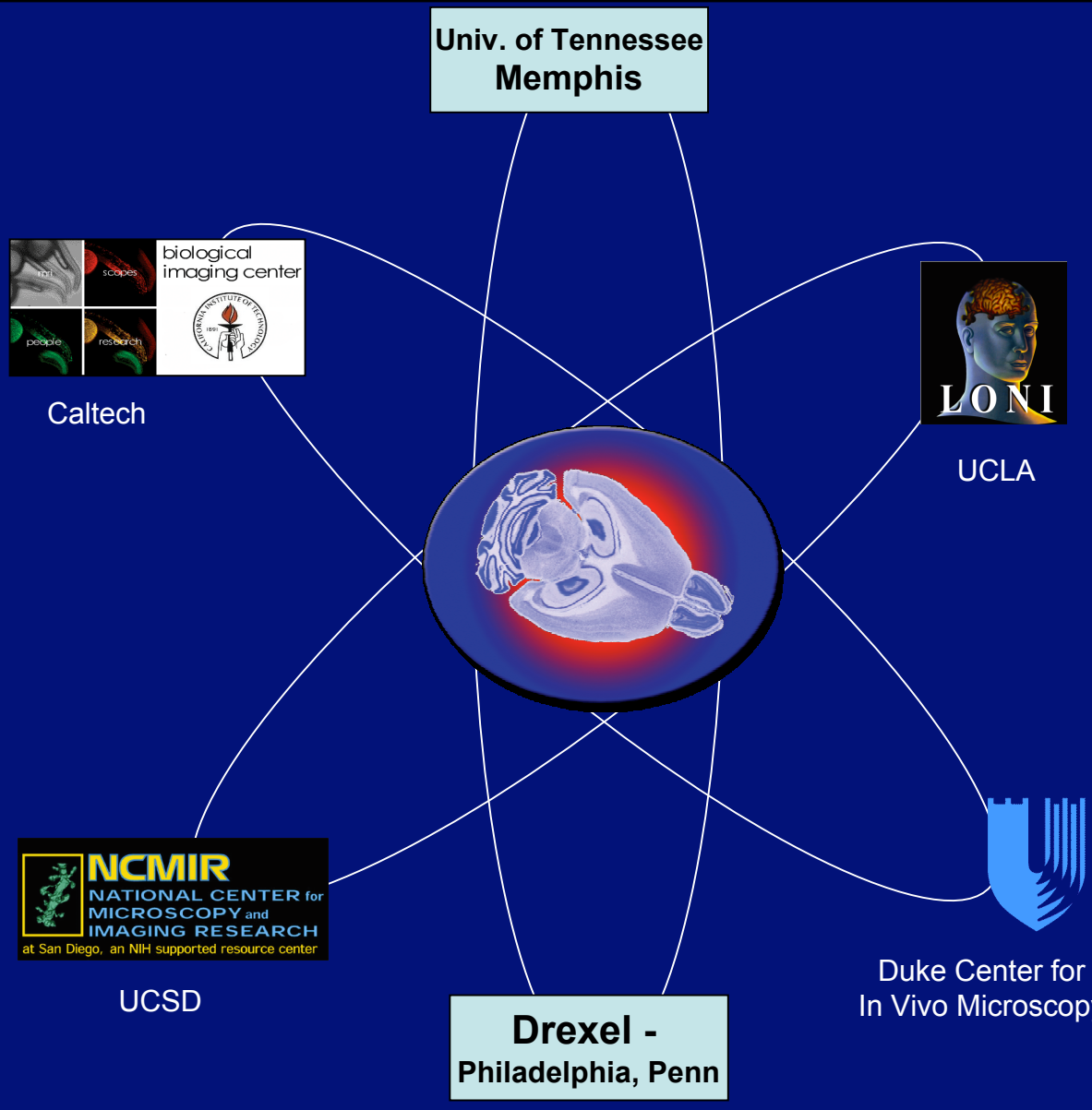
Mouse BIRN

Studying animal models of disease across dimensional scales to test hypothesis with human neurological disorders

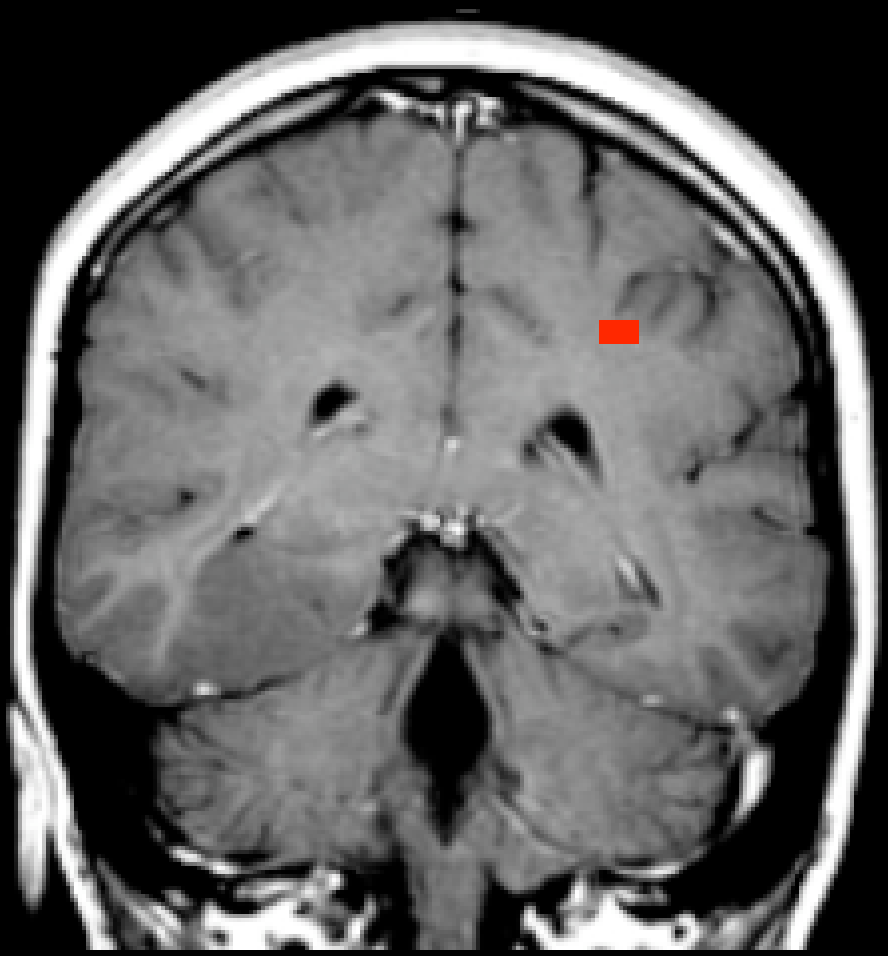
- Experimental Allergic Encephalomyelitis (EAE) mouse models (both chemically induced and transgenic) exhibit episodic weakness and **demyelination** characteristic of **Multiple Sclerosis (MS)**
- Dopamine Transporter (DAT) knockout mouse for studies of **schizophrenia**, **attention-deficit hyperactivity disorder (ADHD)**, Tourette's disorder, and substance abuse
- Using an alpha-synuclein mouse to model the symptoms/pathology of **Parkinson's Disease**
- **Cancer** animal models consortium with astrocytoma mouse model: NCI supported with Terry Van Dyke @ Duke

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

Mouse BIRN



Advanced Imaging - Correlating Human and Mouse



5mm^3

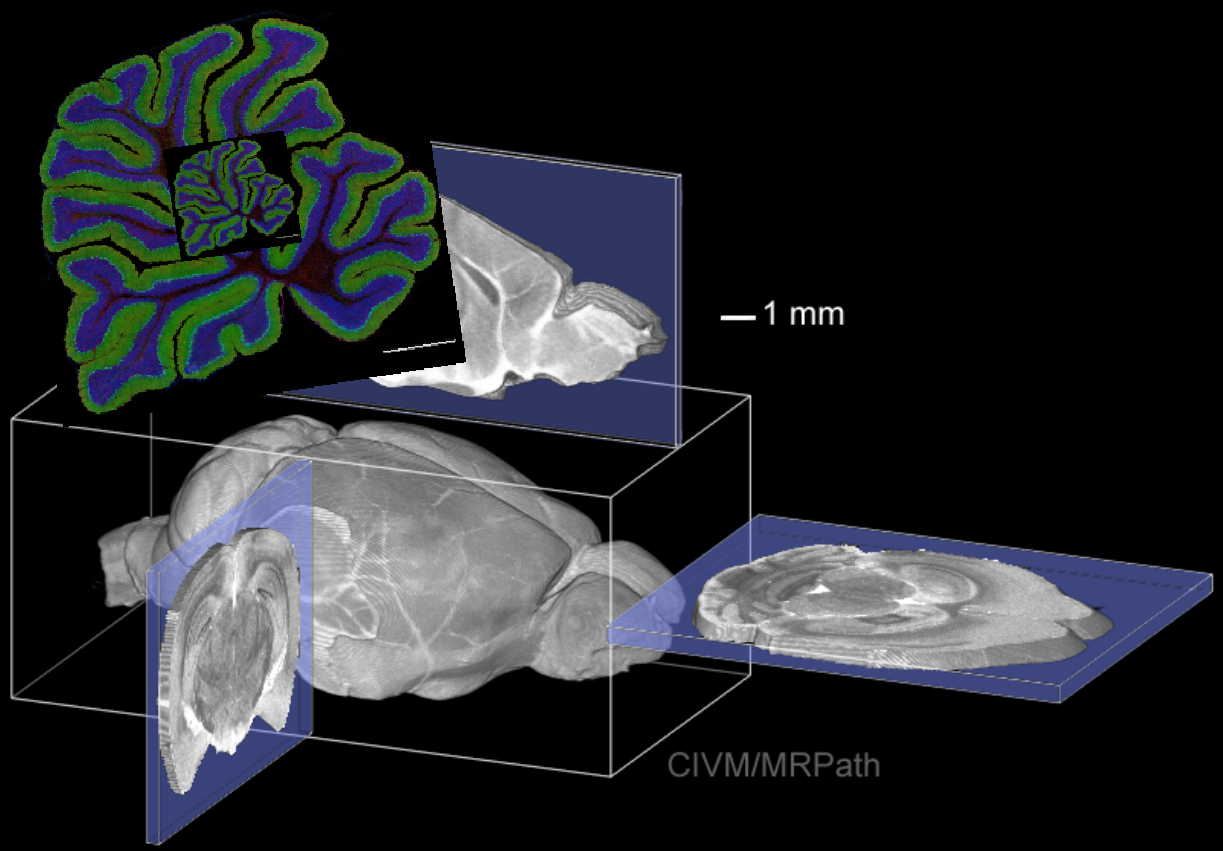
25um^3

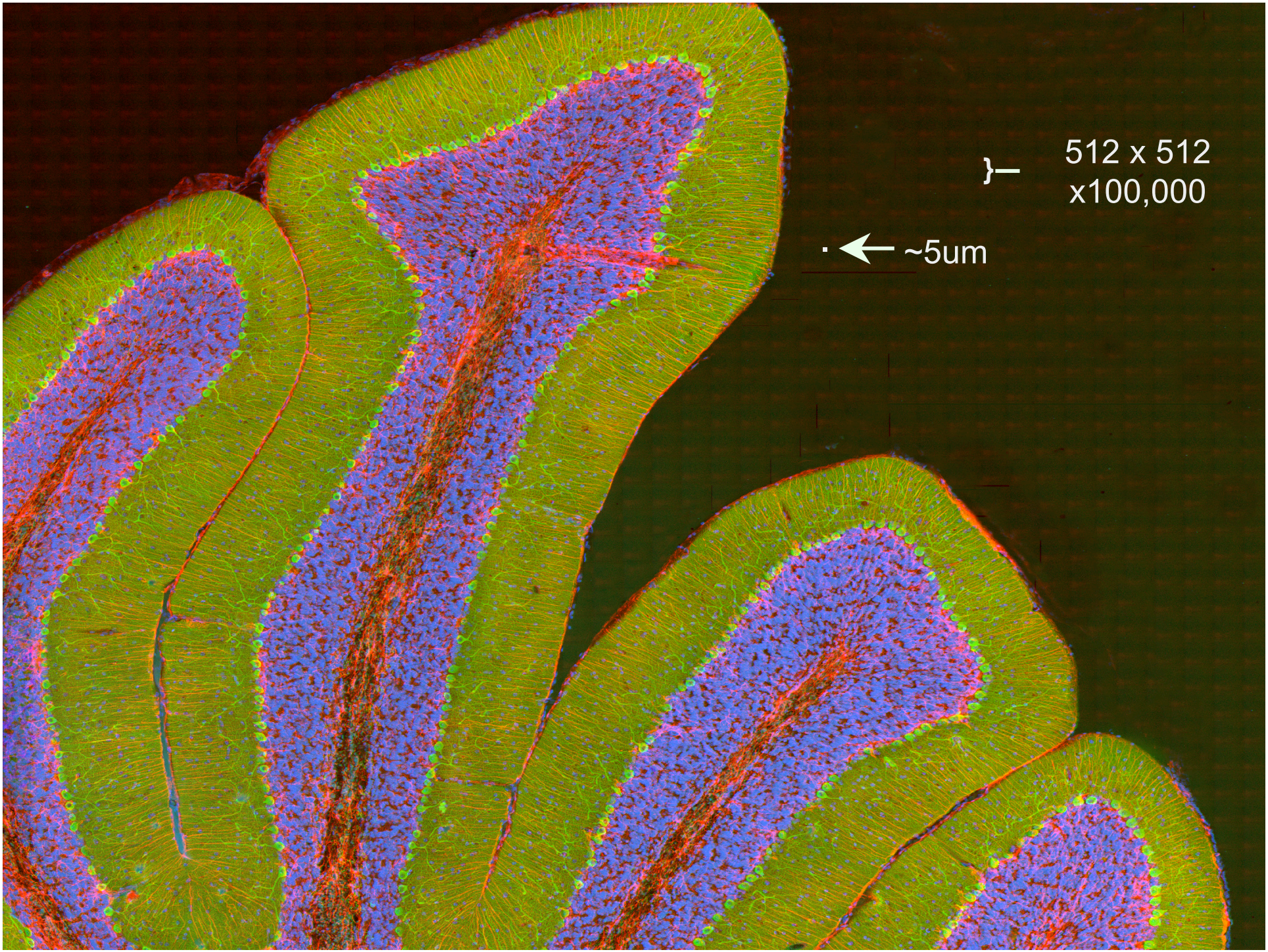
Center for in vivo Microscopy

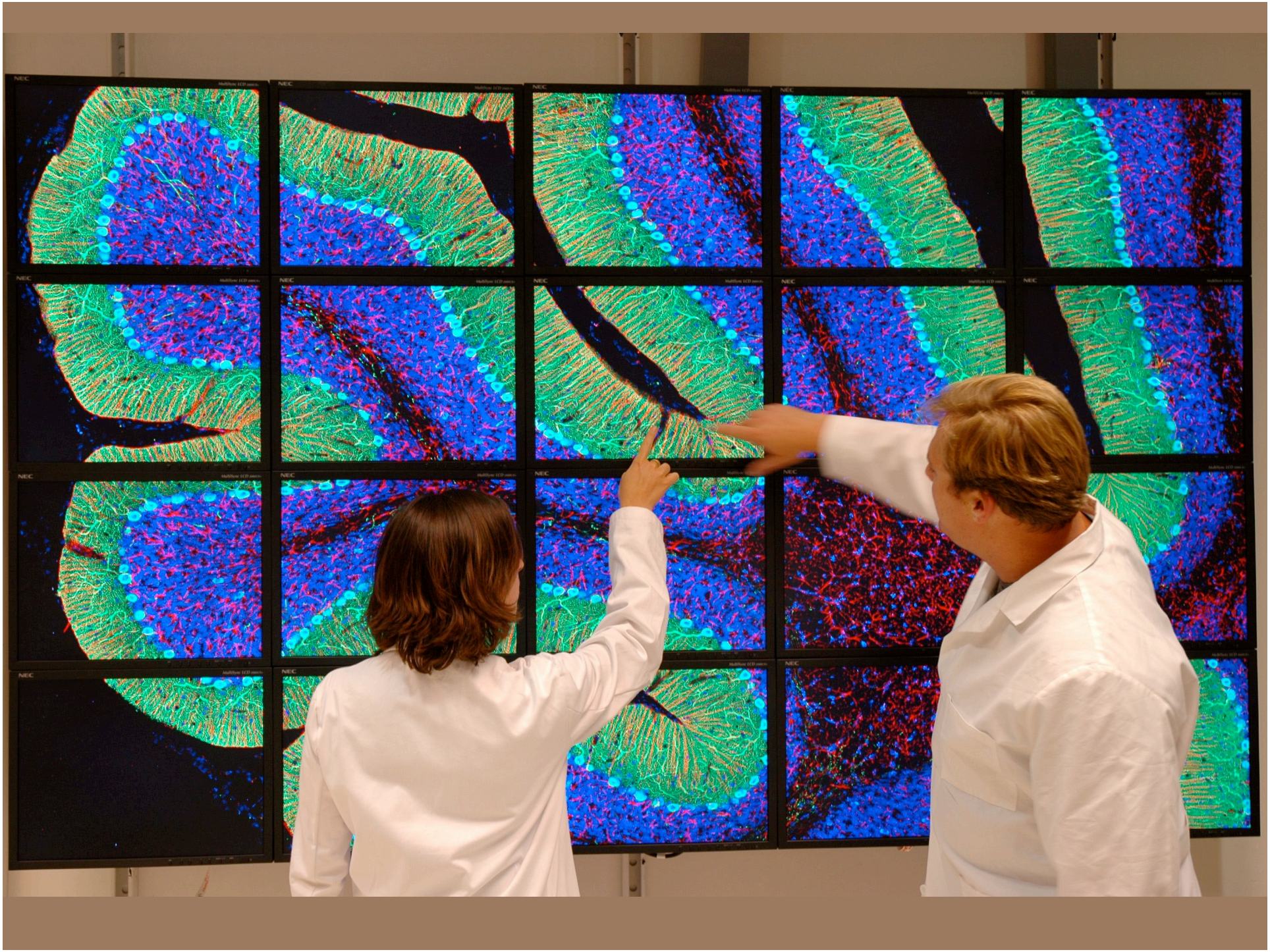
J.A. Johnson - Duke Univ.

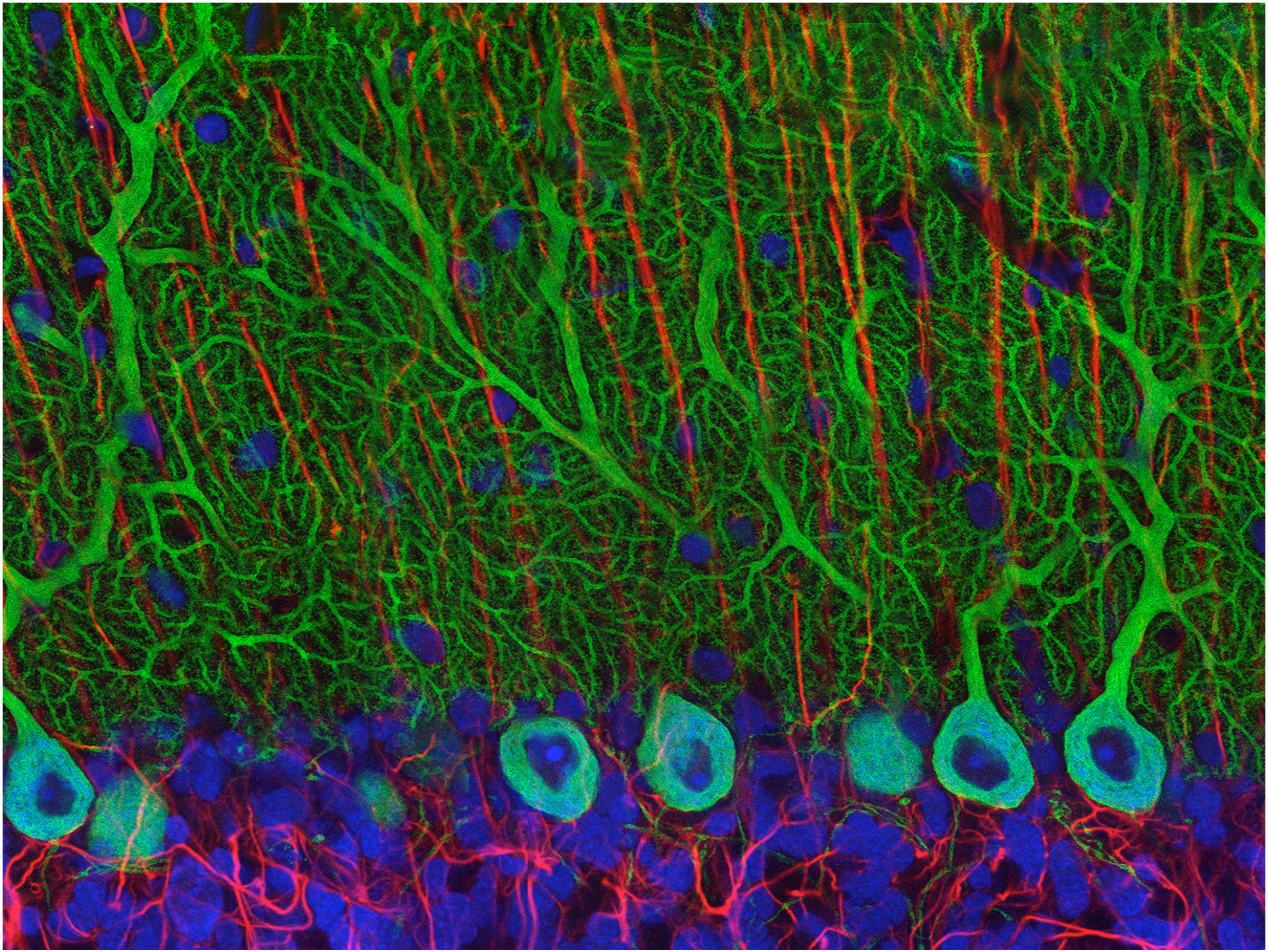
Integration of Multi-resolution data

Mouse BIRN - Duke, Caltech, UCLA, U. Tenn. Memphis, Drexel, UCSD

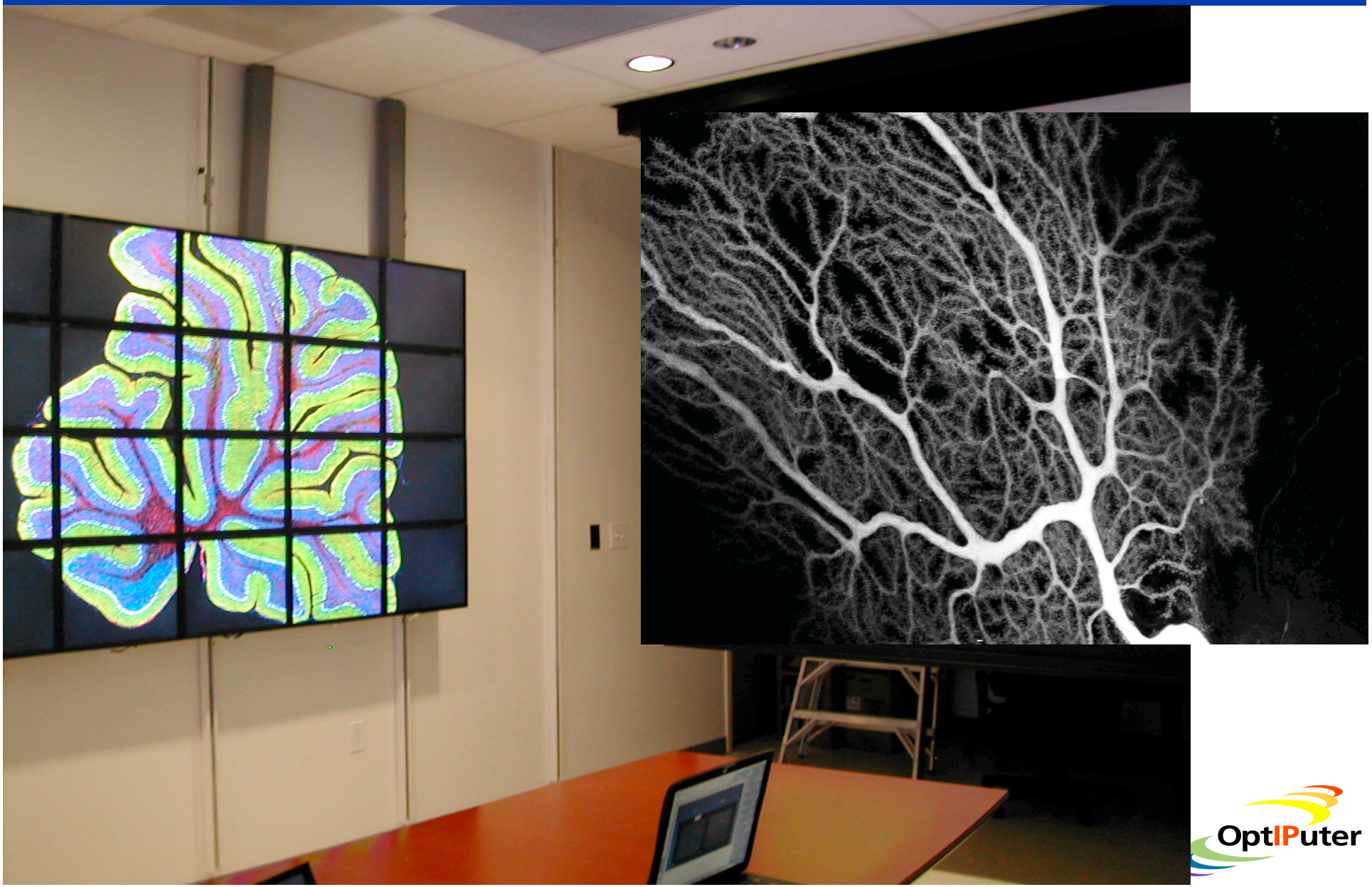



512 x 512
x100,000
~5um

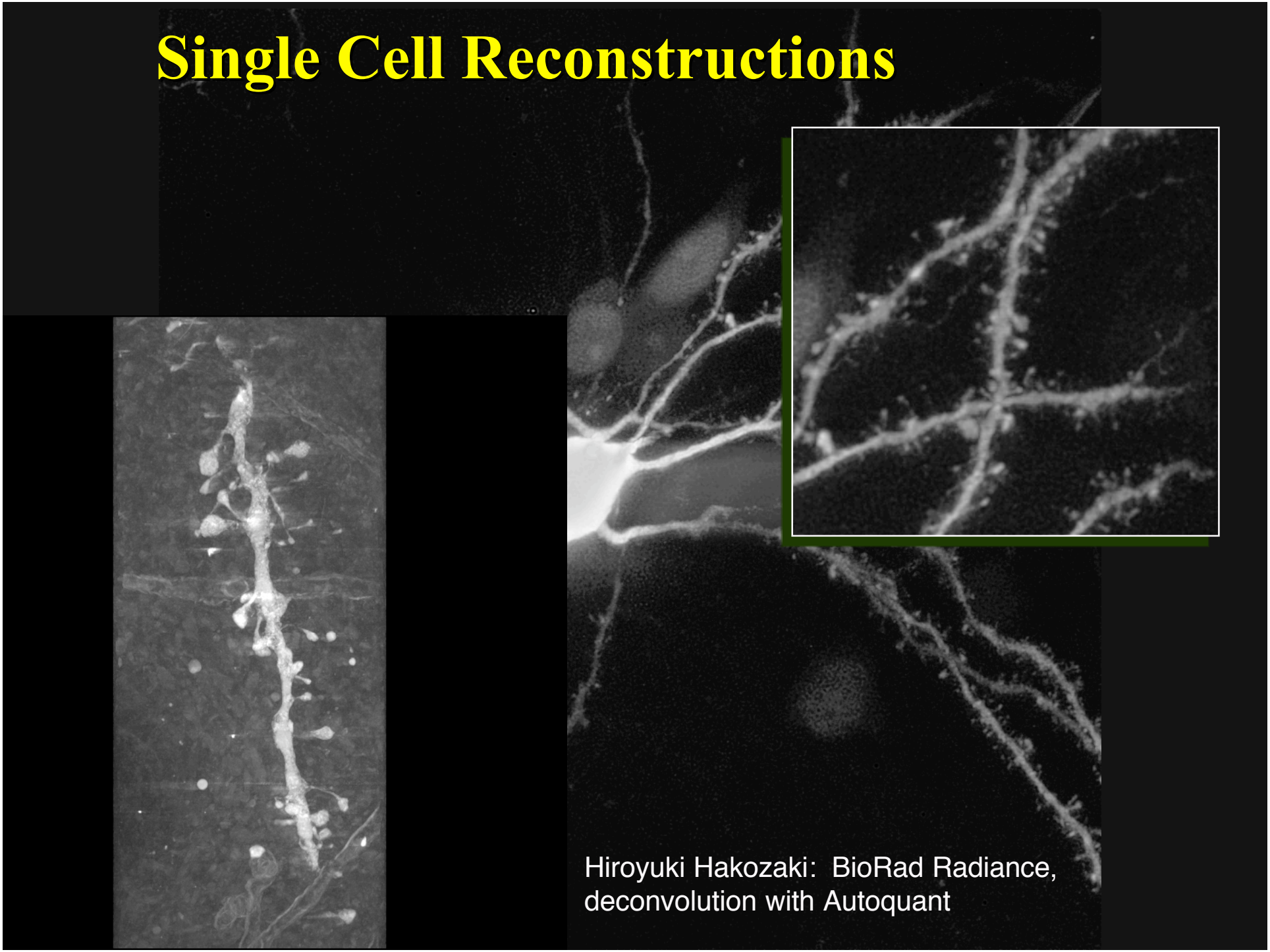




OptIPuter Visualization Environments at NCMIR



Single Cell Reconstructions



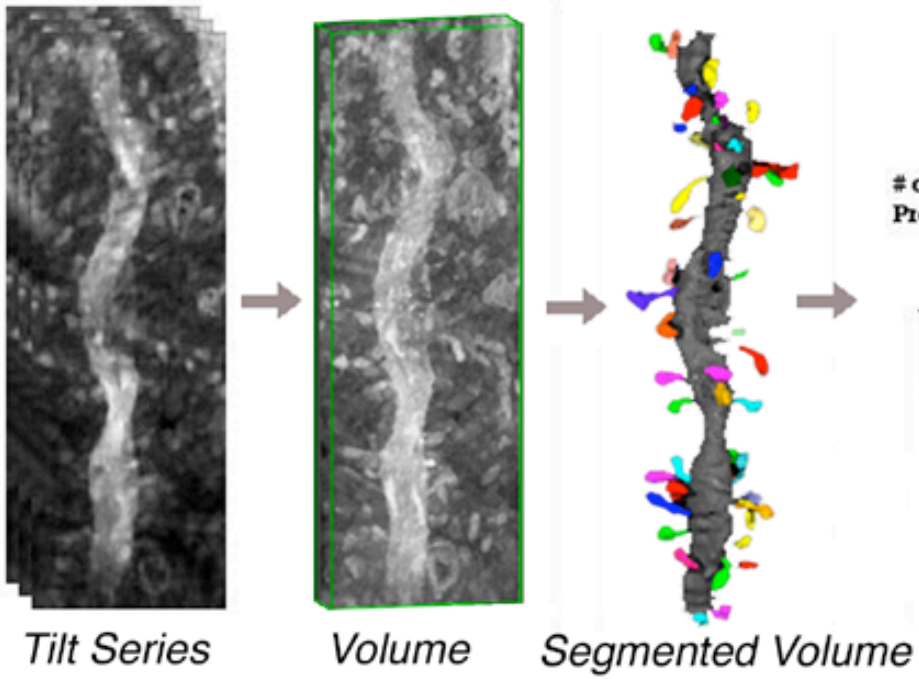
Hiroyuki Hakozaki: BioRad Radiance, deconvolution with Autoquant

NCMIR

NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH

at San Diego, an NIH supported resource center

Data Modeling and Deposition in the Cell Centered Data Base (CCDB)



Tilt Series *Volume* *Segmented Volume*

Segmentation Info

 # objects: 67 **Method**: manual
Program: Xvoxtrace **Segmented by**: M. Martone

Object	Description
sp1	Dendritic spine
sp5	Dendritic spine
sp9	Dendritic spine
sp57xz	Dendritic spine, traced in xz plane
sp62pt	Dendritic spine, partial

Measurements

	SA	Vol	Length	#
sp1	1.1	0.05	1.2	1
sp5	0.4	0.01	0.8	1
sp9	1.53	0.05	1.0	1
sp57xz	1.36	0.05	1.2	1
sp62pt	na	na	na	na

Units=μm

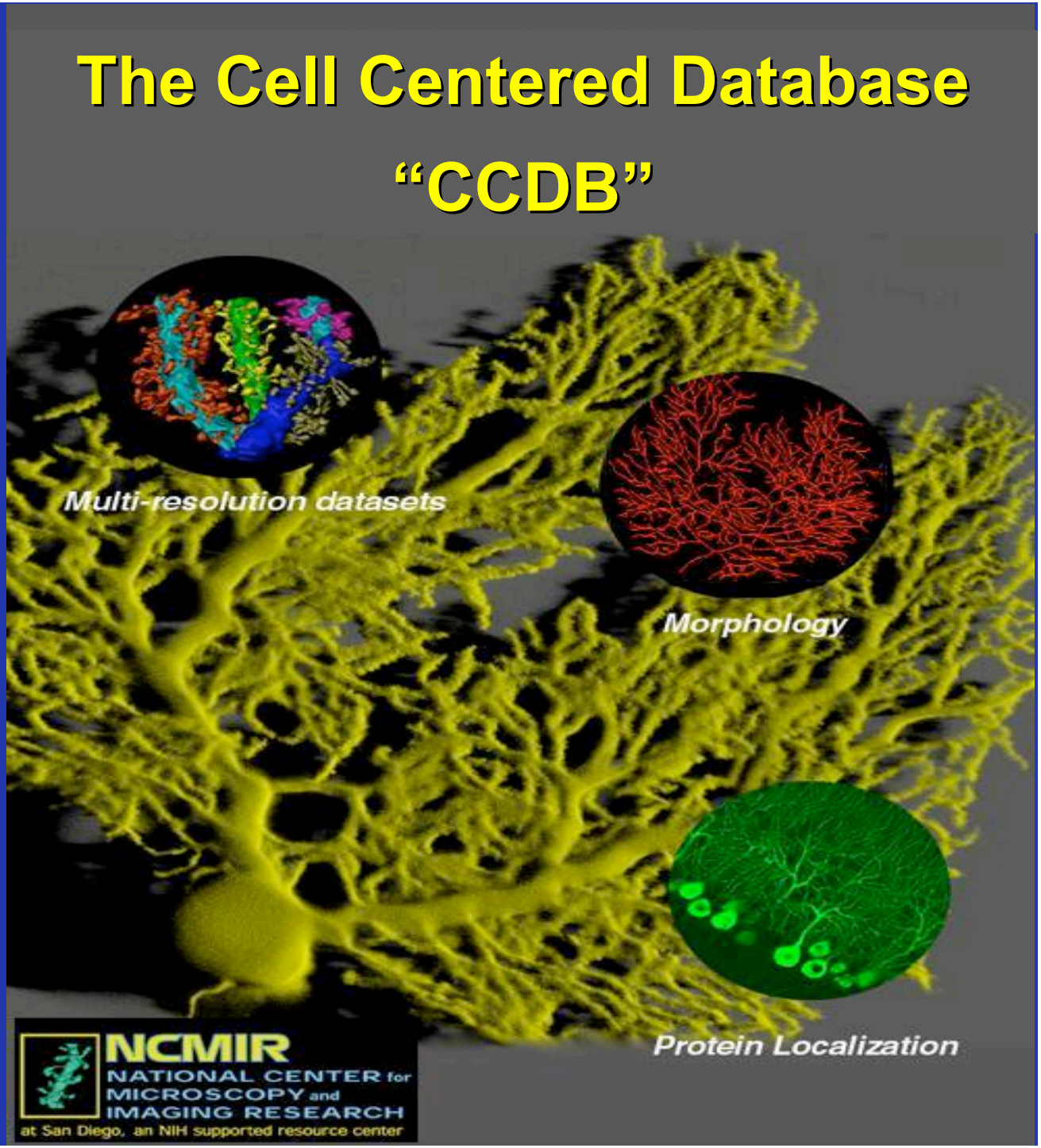
Measurement Info

Measurement	Programs Used
SA	Analyze
Vol	Analyze
Length	Xdend
Number	Manual

from Maryann Martone, Amarnath Gupta, Bertram Ludaescher, Naoko Yamada and Mona Wong

The Cell Centered Database "CCDB"

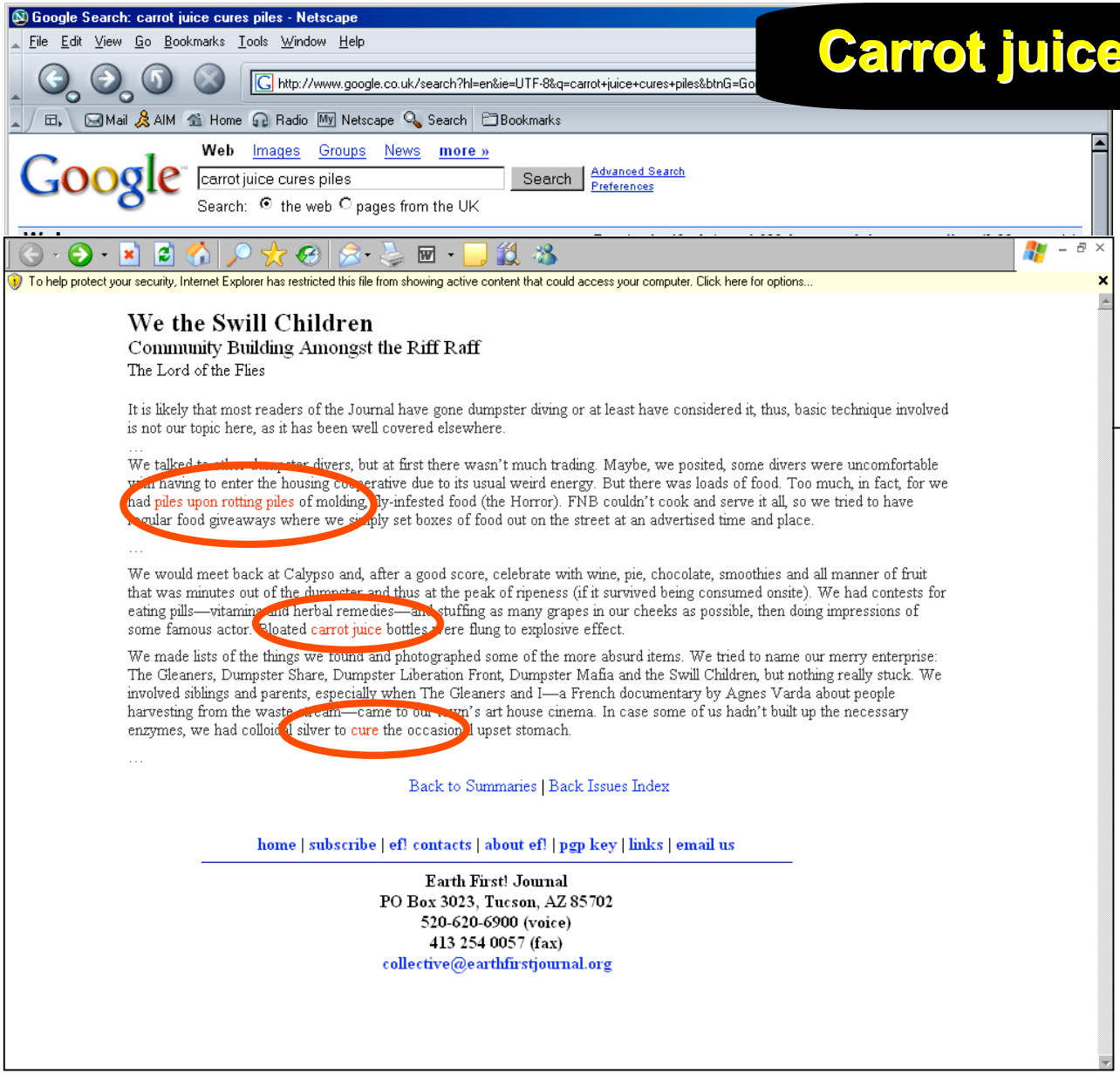
- A Federated Distributed Database for Neuroscience
- A Multimode & Multiscale "DataGrid"
- Interoperates with Gene and Protein databases & "brain map" databases of brain anatomy



From:
Maryann Martone, Amarnath Gupta, Bertram Ludaescher, Naoko Yamada, Yujun Wang, Julia Sun, and Mona Wong

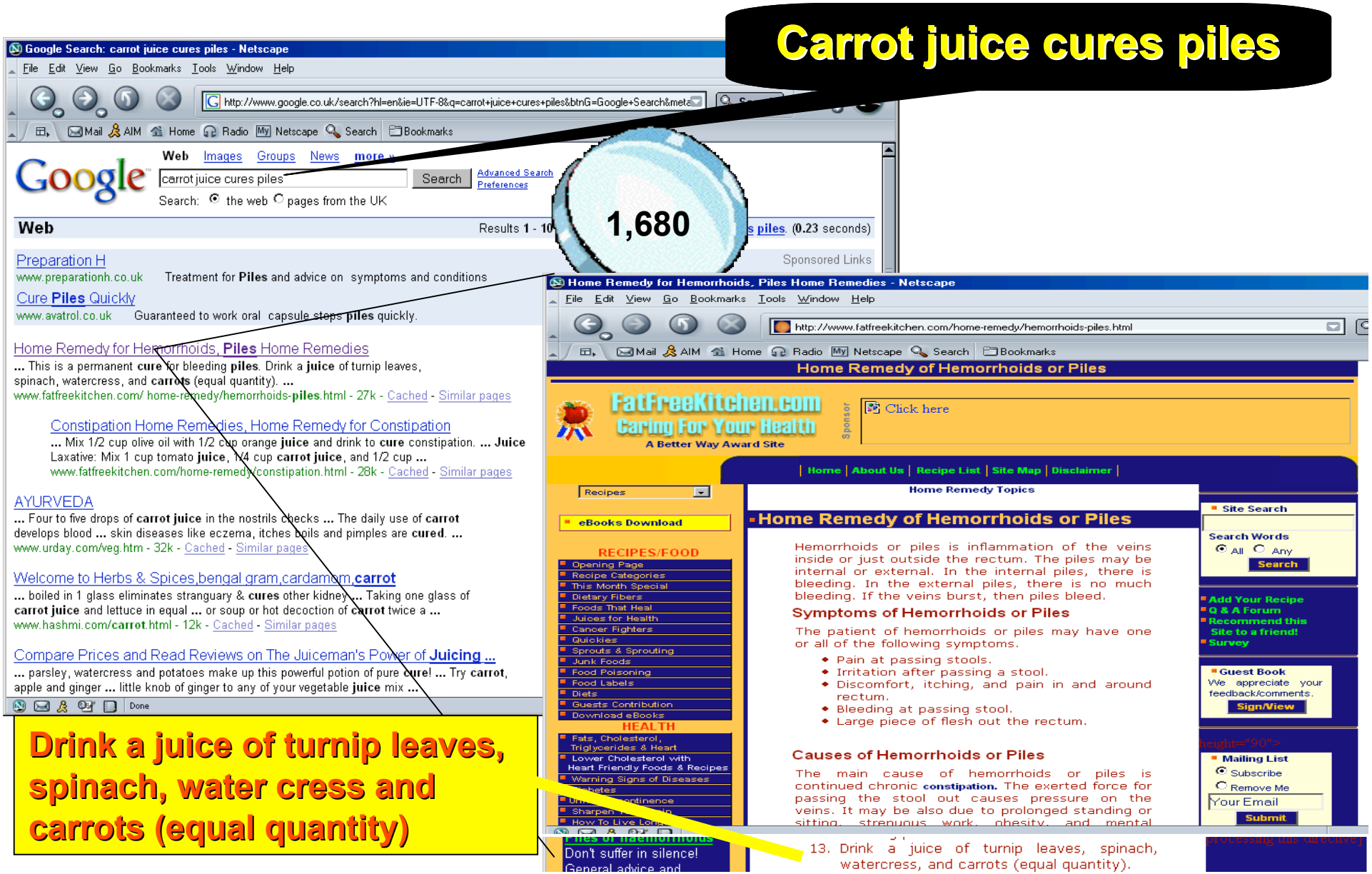
Google is not a portal

Carrot juice cures piles

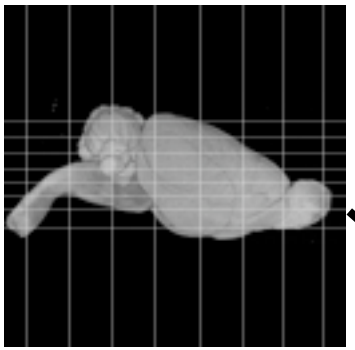


What Ken Peach expected

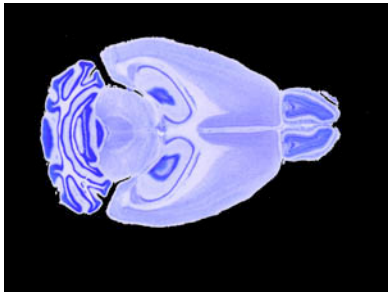
What he did not expect



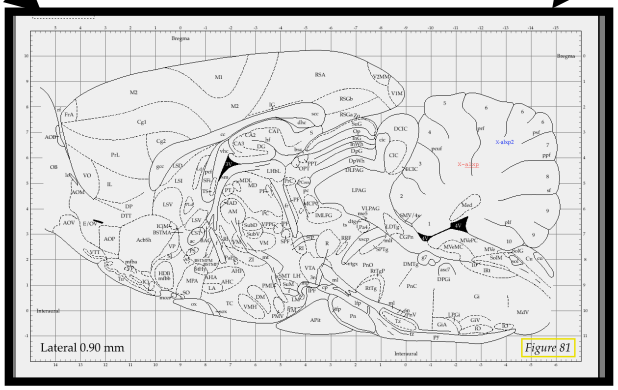
BIRN Data Integration: *An example of federation of Multi-scale, Multi-modal data from Mouse BIRN*



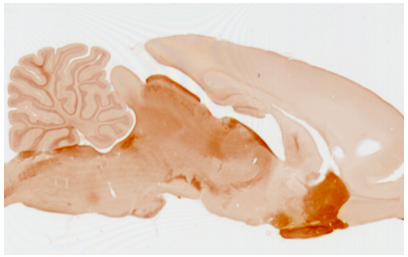
1. Create databases at each site



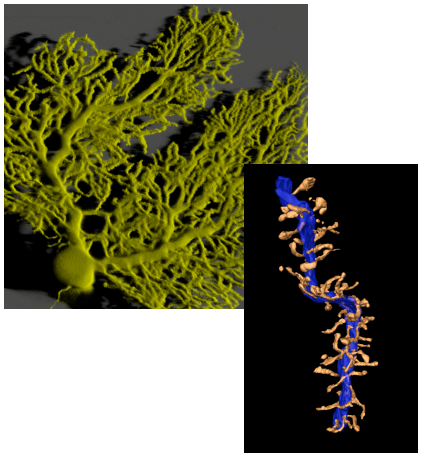
2. Create conceptual links to a shared **"ONTOLOGY"**



4. Use **integration engine "Mediator"** to navigate and query across data sources



3. Situate the data in a **common spatial framework**

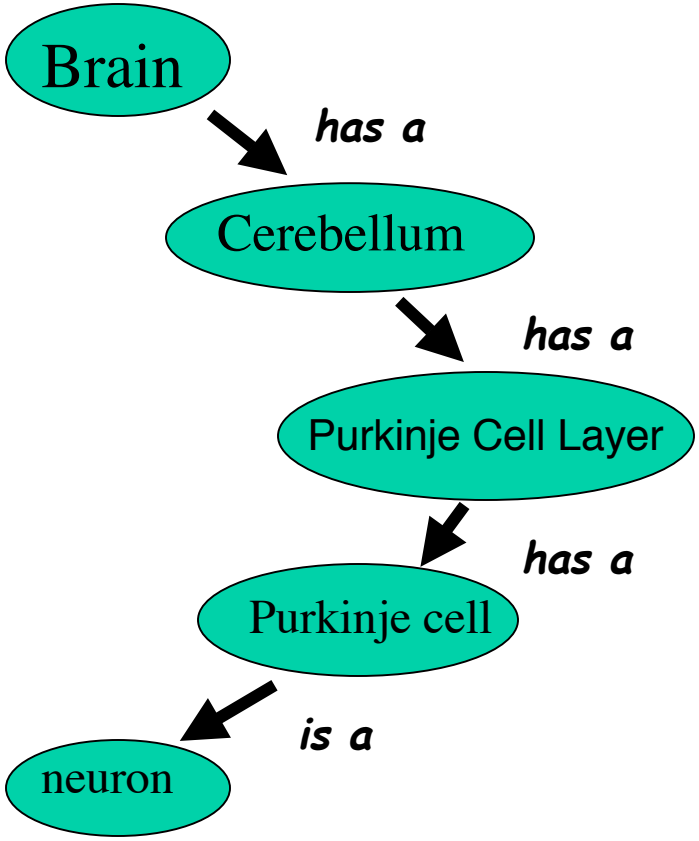


DATA MODELS: Frameworks to Integrate Databases

BIRN uses "Integrated Views" based on "Ontologies"

What is an **Ontology**?

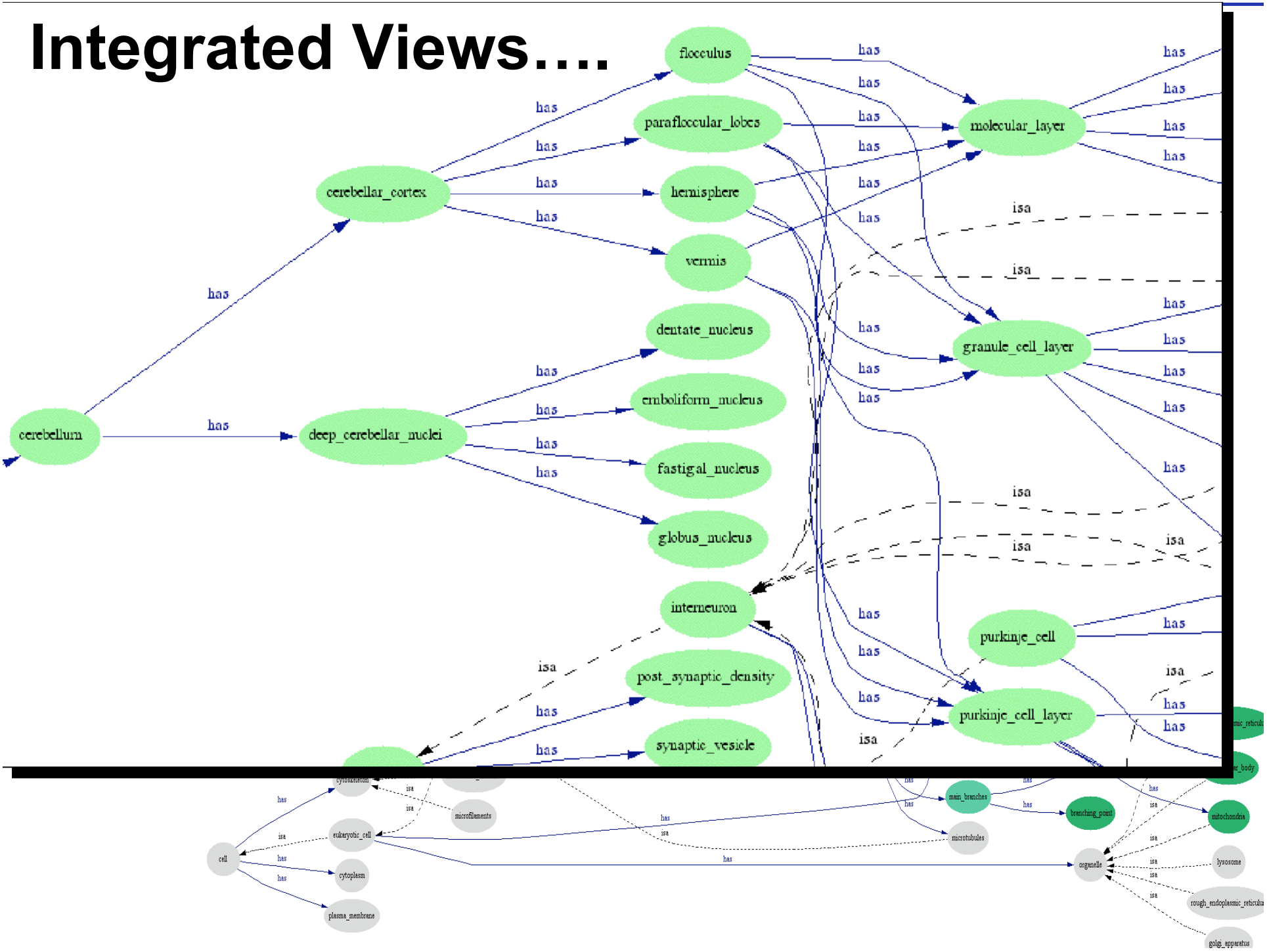
- a way to represent and communicate a shared understanding of a field
- representation of terminological knowledge
- explicit specification of a conceptualization
- concept hierarchy ("***is-a***")
- further semantic relationships between concepts ("is part of", "causes", etc.)



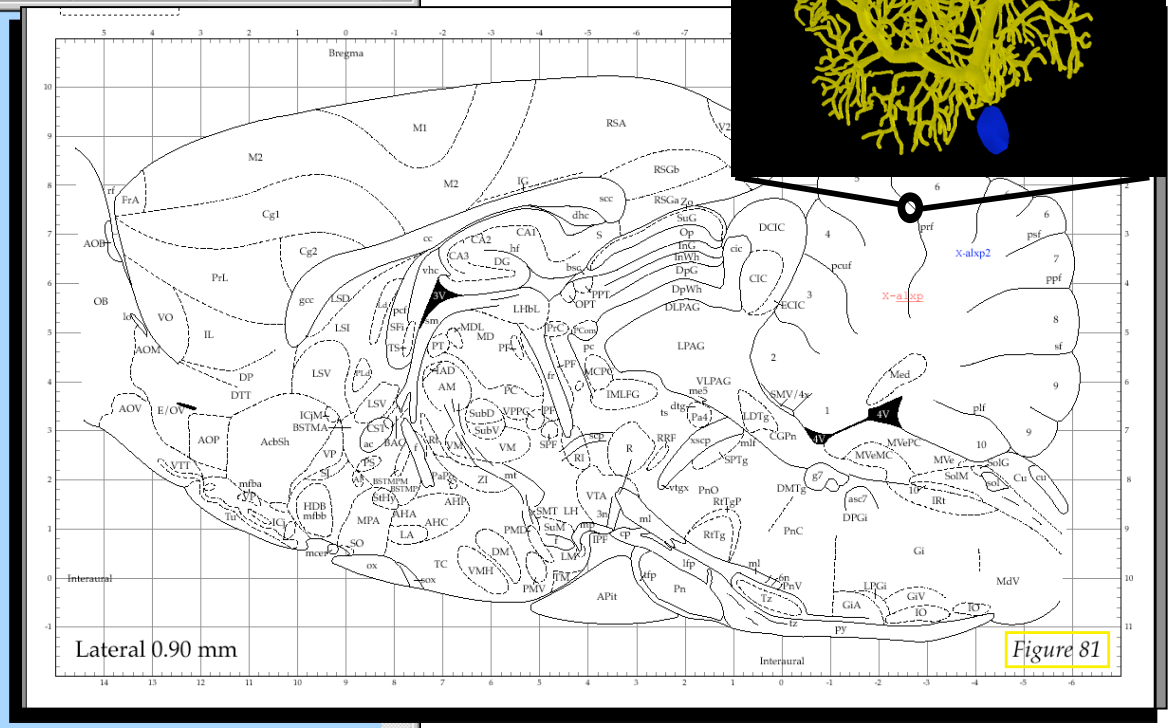
Examples:

- NCMIR ANATOM
- GO (Gene Ontology)
- UMLS (Unified Medical Language System)
- CYC

Integrated Views….



Atlas-based Spatial Reference Systems: *Multi-Scale and Multi-Modal Data are Connected through Ontologies*



Update user-defined area -

https://pamina2.sdsc.edu/rat-atlas/Scripts/UpdatePolygon.php?uid=Josh

Geometry name:
cerebellum

Attached data:

URL	http://www.myurl.com
File path	
Database Name	ccdb
Database ID	33

More boxes

Annotation:

Purkinje neuron

Save

Close

Done

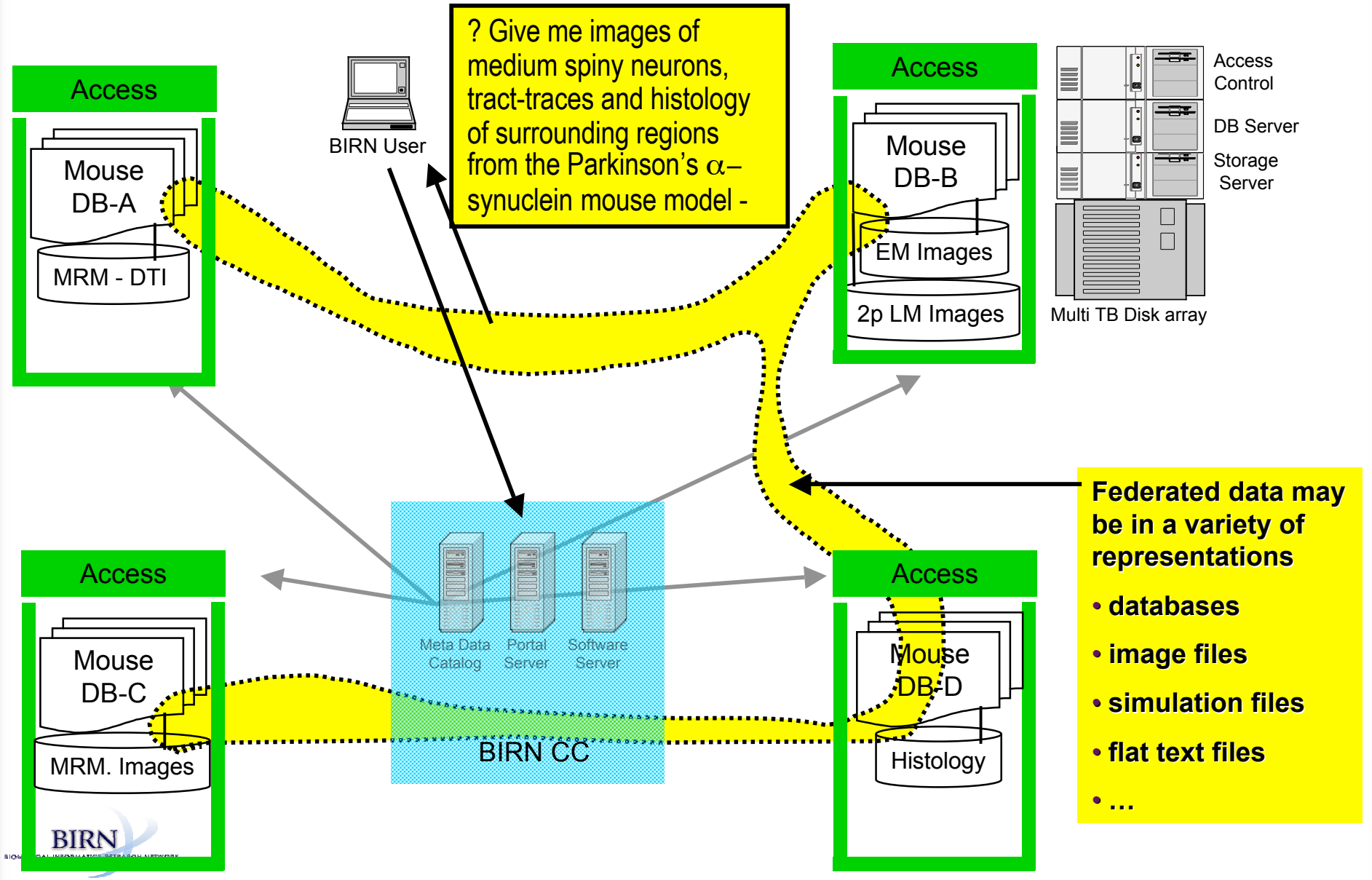
UMLS ID

Registering My Data

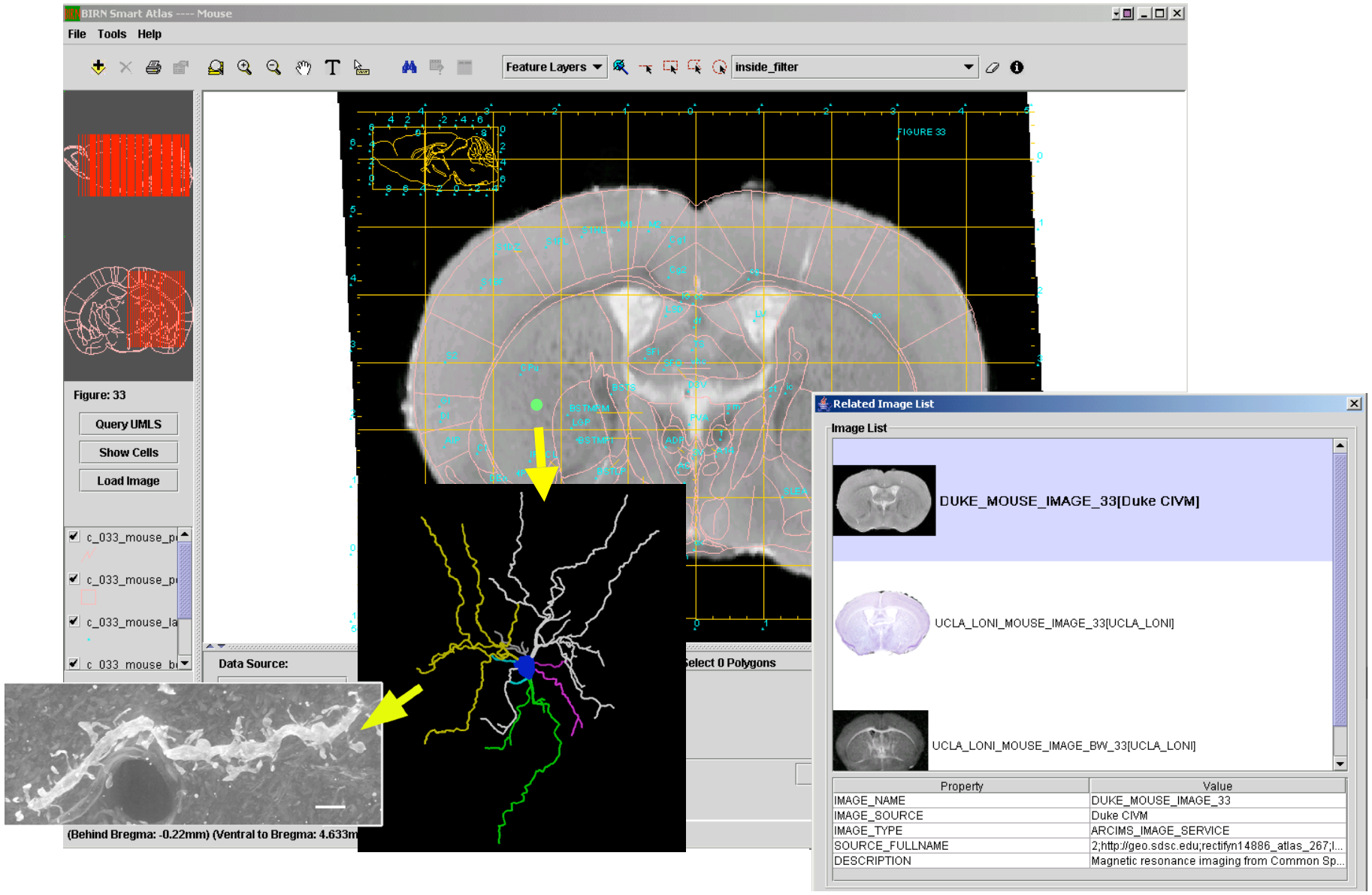
or use "KNOW ME" tool

Location, Location and Location…….. and a Unique Identifier

BIRN Data Integration Environment Bridges Data Models When Users Explore Distributed Data

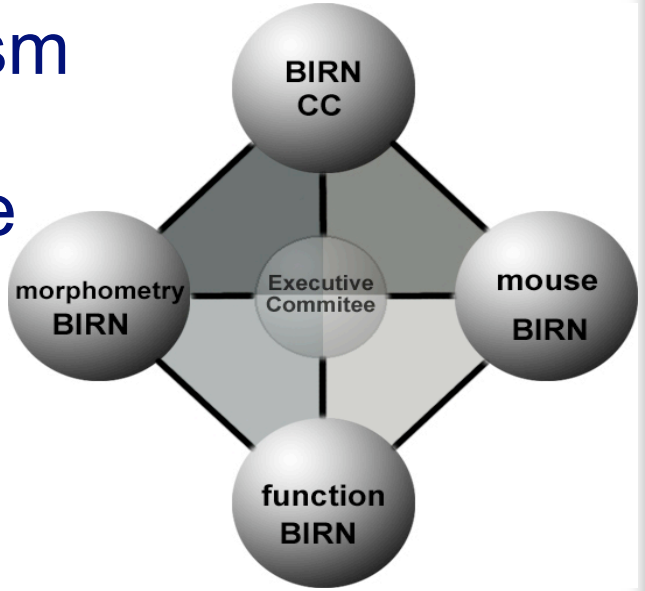


The BIRN Smart Atlas: *An Example of a Data Grid-based GIS-like tool for spatial integration of multiscale distributed brain data. *Runs from BIRN Portal*



BIRN is a Growing Consortium

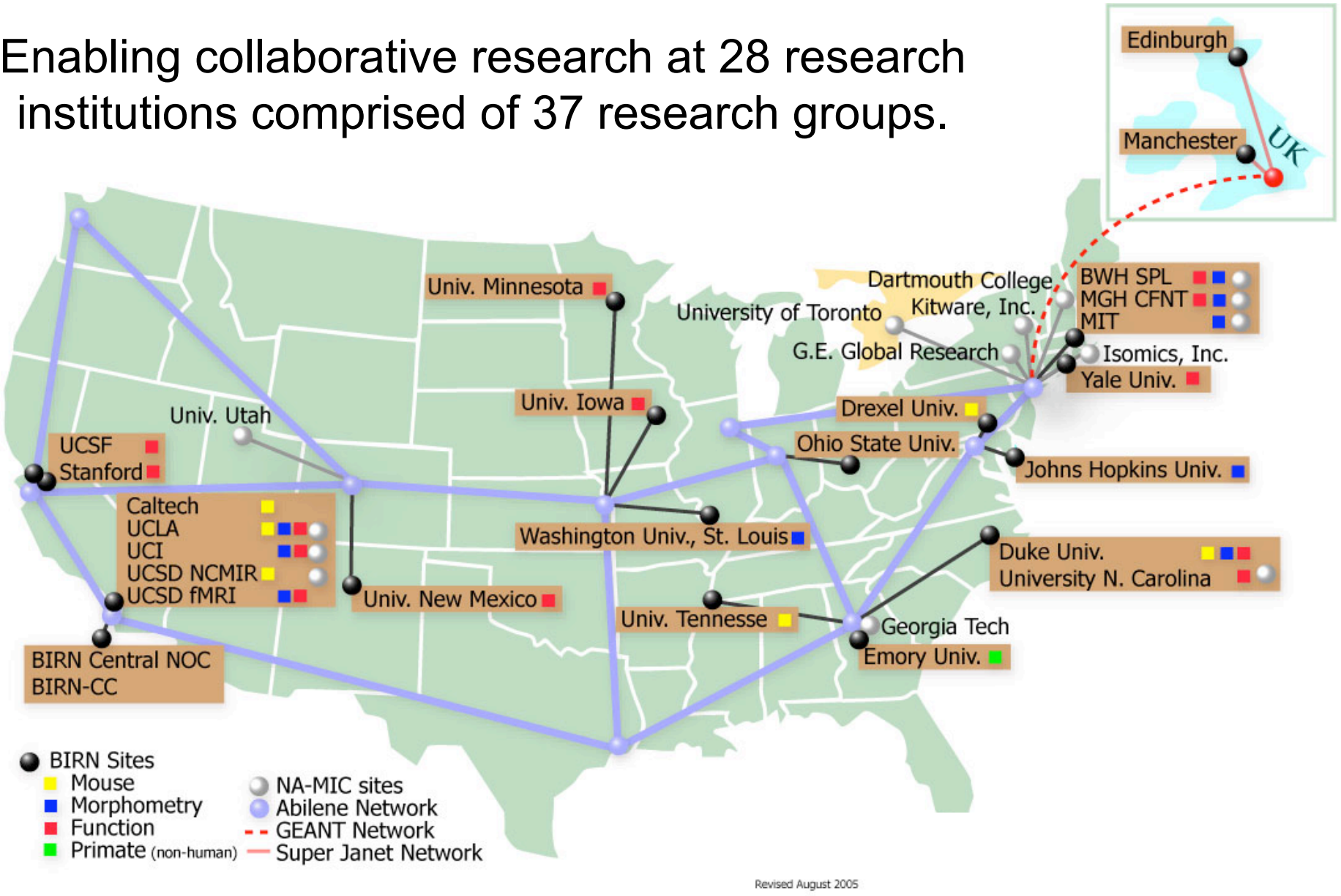
- NDAR - National Database for Autism Research for the support of clinical and translational research within the autism research community
- Non-Human Primate - Linking imaging, behavior, and molecular informatics in non-human primate pre-clinical models of disease
- UK e-Science - Fostering collaborations with European cyberinfrastructure community



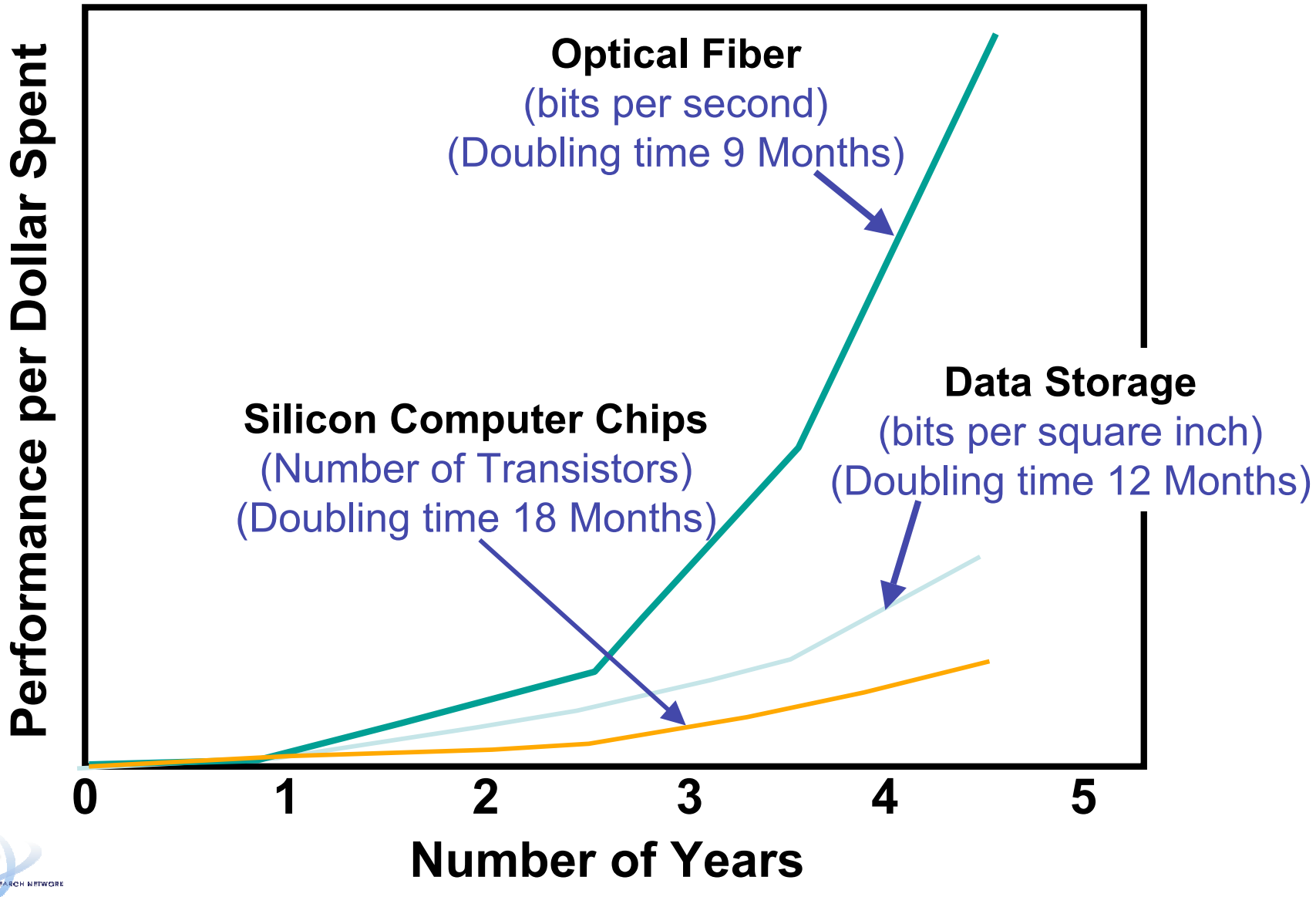
BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

The BIRN Collaboratory Today

Enabling collaborative research at 28 research institutions comprised of 37 research groups.

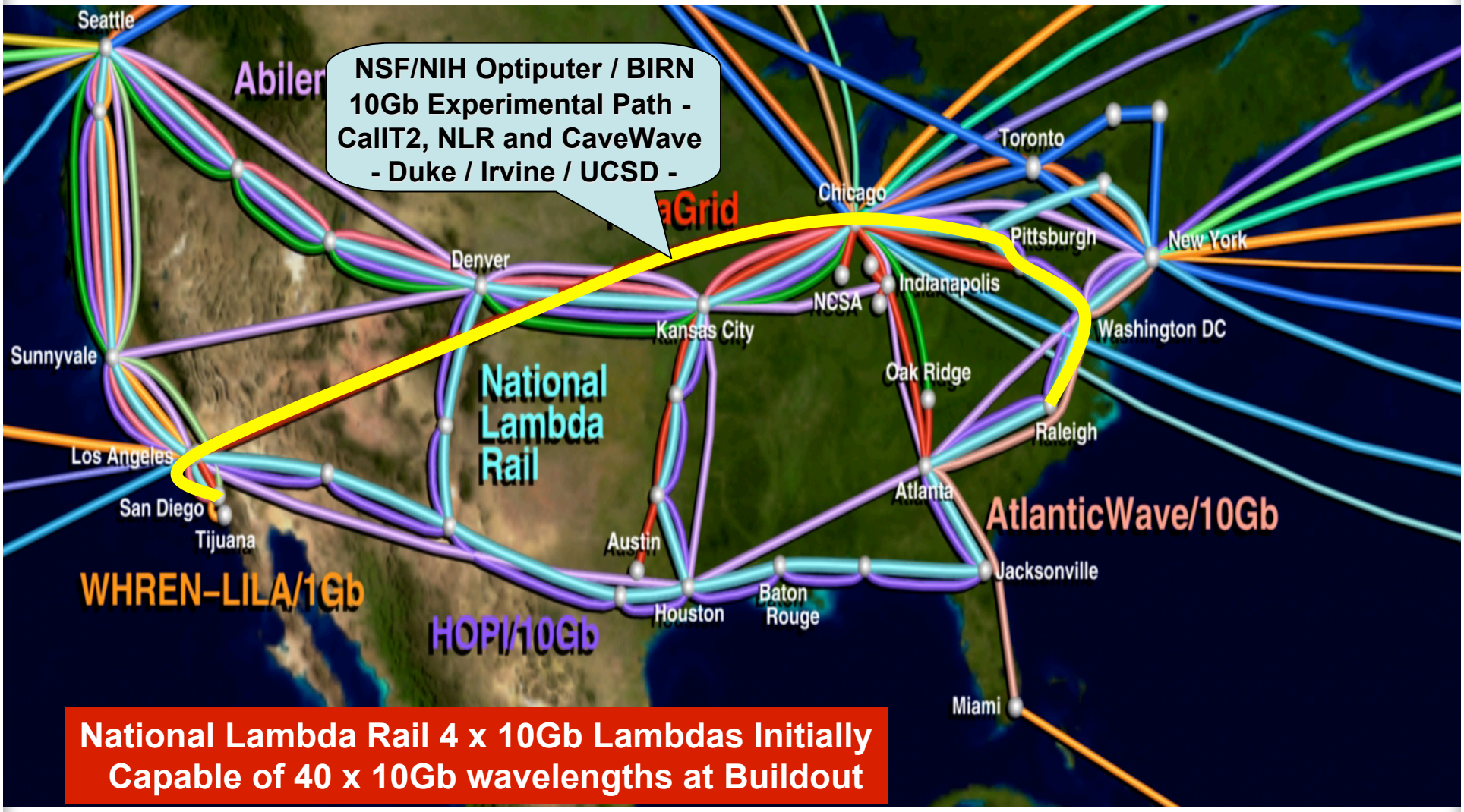


Optical Networks Are Becoming the 21st Century Cyberinfrastructure Driver



Scientific American, January 2001

Building Bridges: NIH Research and Infrastructure Projects & NSF Advanced Network Research and Infrastructure Projects



BIRN BIOMEDICAL INFORMATICS RESEARCH NETWORK

Centered around anticipated research requirements

OptIPuter

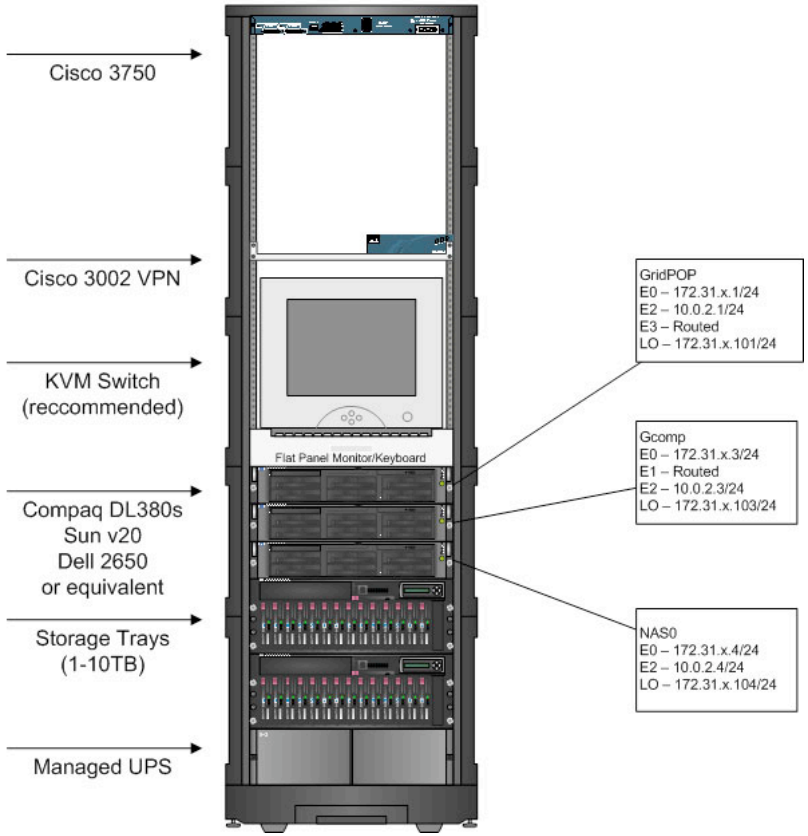
Removing Barriers to BIRN:

Decreasing Cost of Entry & Increasing Scalability

$120K
(2001)

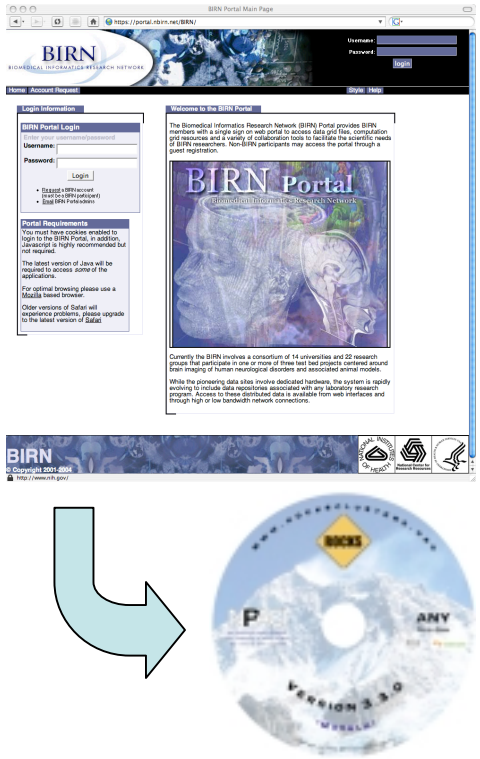
- Prescribed hardware jumpstarted BIRN for functionality

< $20K
(Today)



- Support for multiple vendors

< $5K
(~2011)



- Software solution for researchers to BIRN "enable" local hardware

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

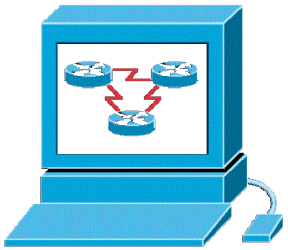
BIRN is a Stable & Rapidly Evolving National Research Infrastructure

- ✓ Supporting collaborative activities of advanced biomedical research & clinical research centers in the US - Serving as a model for programs everywhere.
- ✓ Ensuring a stable, robust, shared network environment across > 35 institutions today - High Bandwidth Connectivity via Internet2.
- ✓ Developing hardware and software infrastructure for managing distributed data - creation of the BIRN Data GRID.
- ✓ Providing secure and audited access to distributed data - deployment of a Uniform BIRN Security Model.
- ✓ Exploring data using "intelligent" query engines that can make inferences upon locating "interesting" data - development of the BIRN Data Integration Environment.
- ✓ Integrating BIRN with middleware projects in academia & industry - facilitating the use of Computational GRID infrastructure.
- ✓ Providing simple and intuitive access to a shared processing, visualization and analysis environment - BIRN is a leader in GRID Portal technology.
- ✓ Changing the use pattern for research data from the individual laboratory/project to shared use.
- ✓ Promoting large-scale collaboration among research scientists across institutional boundaries

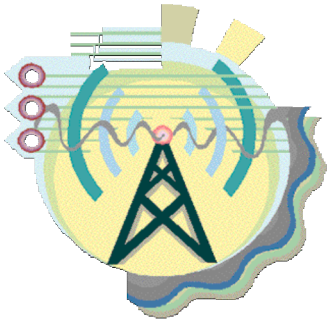
The BIRN Coordinating Center is Supporting and Evolving the Deployed Infrastructure

HELP DESK

MONITORING

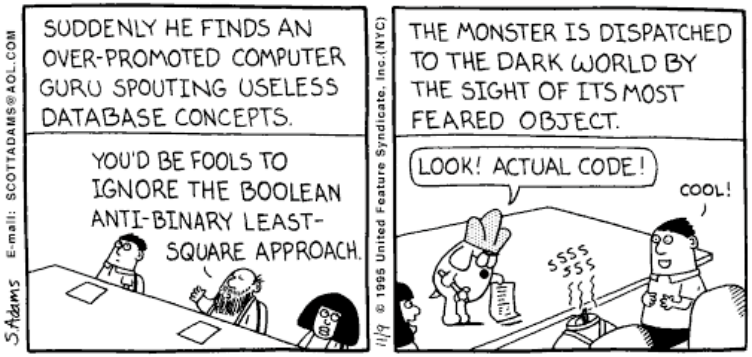


NETWORKING



COMMUNICATIONS INFRASTRUCTURE

TRAINING



ASSESSING TECHNICAL OPTIONS FOR BIRN TO BUILD WORKING SYSTEMS

SCALABLE SOFTWARE DISTRIBUTION

BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

5th Annual
BIRN All-Hands Meeting

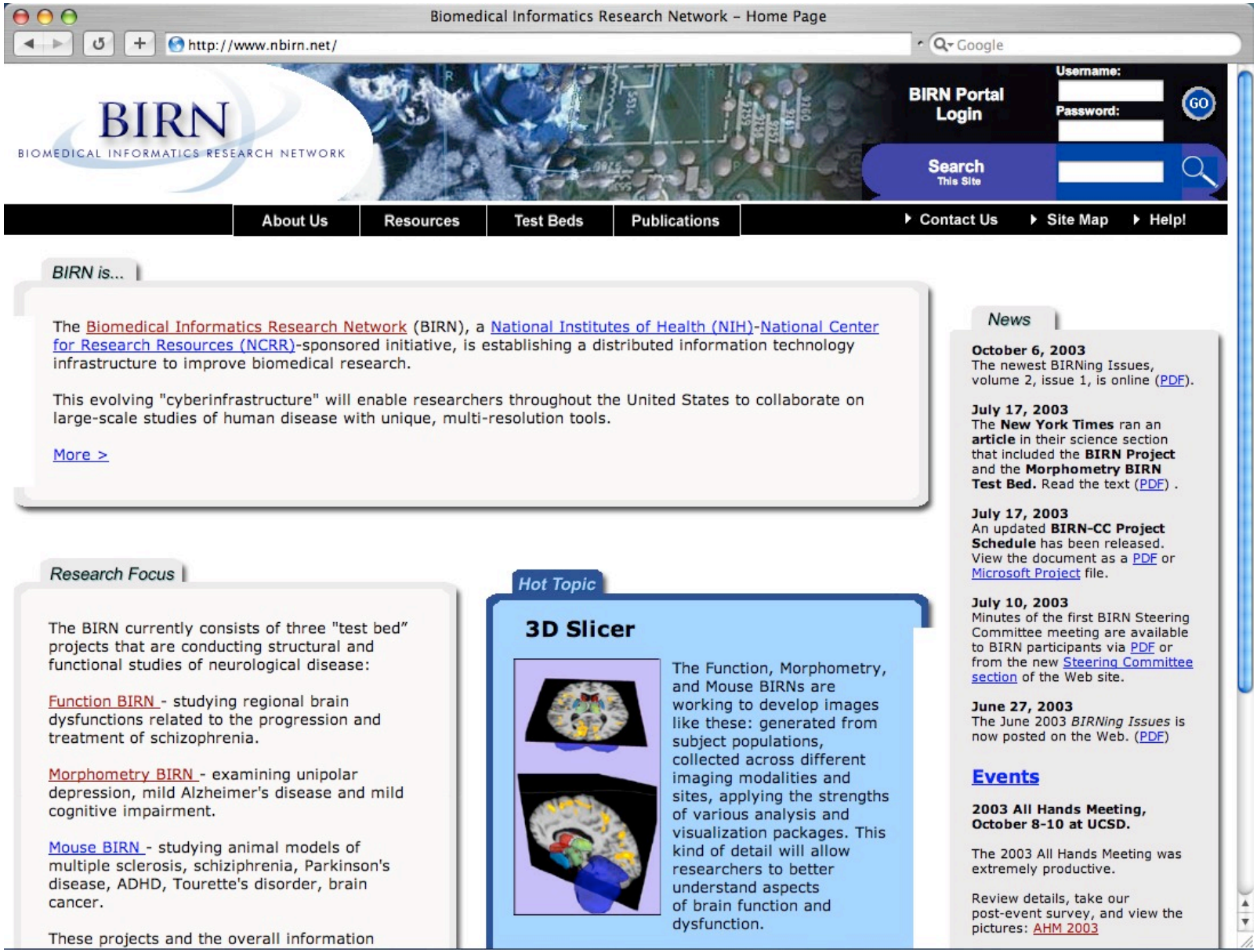
More than 200 BIRN members from across the nation participated in presentations, brain-storming workshops, and problem-solving discussions at the BIRN All Hands Meeting.

UCSD

October 2005
La Jolla, CA

it²

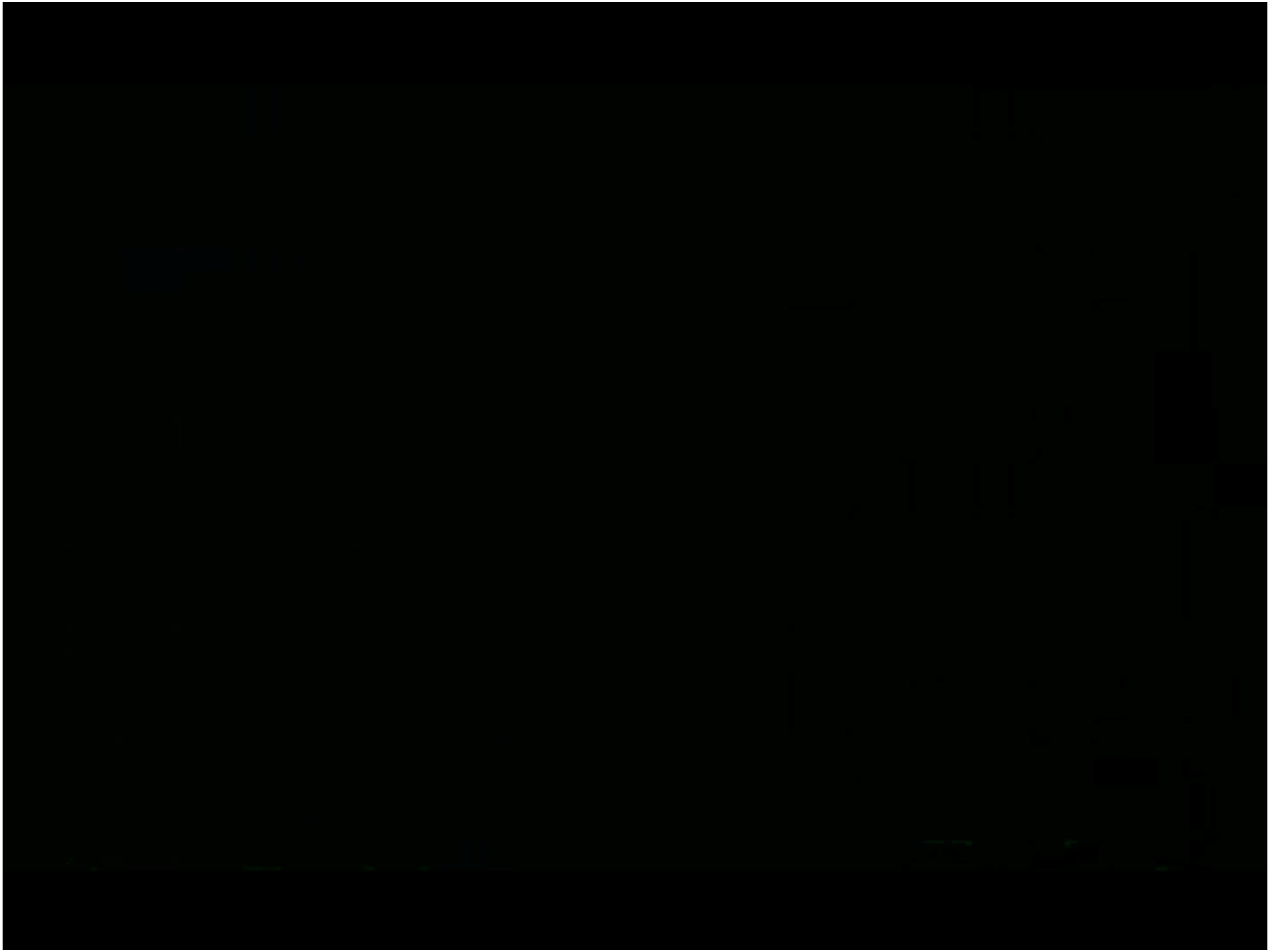
http://www.nbirn.net

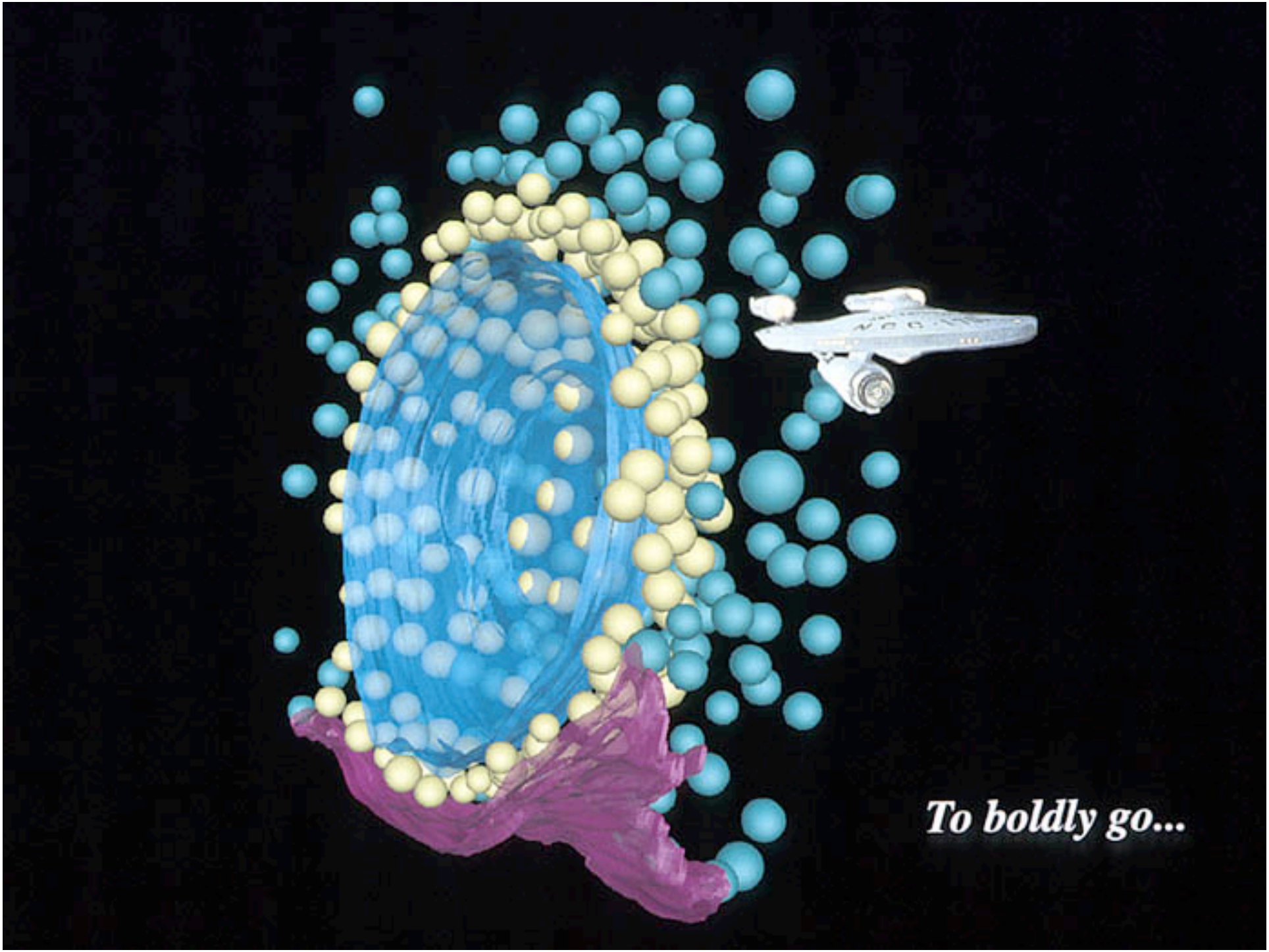


BIRN
BIOMEDICAL INFORMATICS RESEARCH NETWORK

Challenges

- Breaking down the barriers
  - Mistrust
  - Open sharing of information
  - Who gets credit
  - Commercial products
  - Governance
- Incorporating processes for multi-site studies and sharing of human data
  - HIPPA Compliance
  - Patient confidentiality
  - Institutional Review Board (IRB) approvals
- Developing guidelines - for sharing data & authorship
- Integrating new participants
- Providing an architecture to allow for technology improvements with the existing infrastructure
- Guaranteeing security versus ease of use




To boldly go...