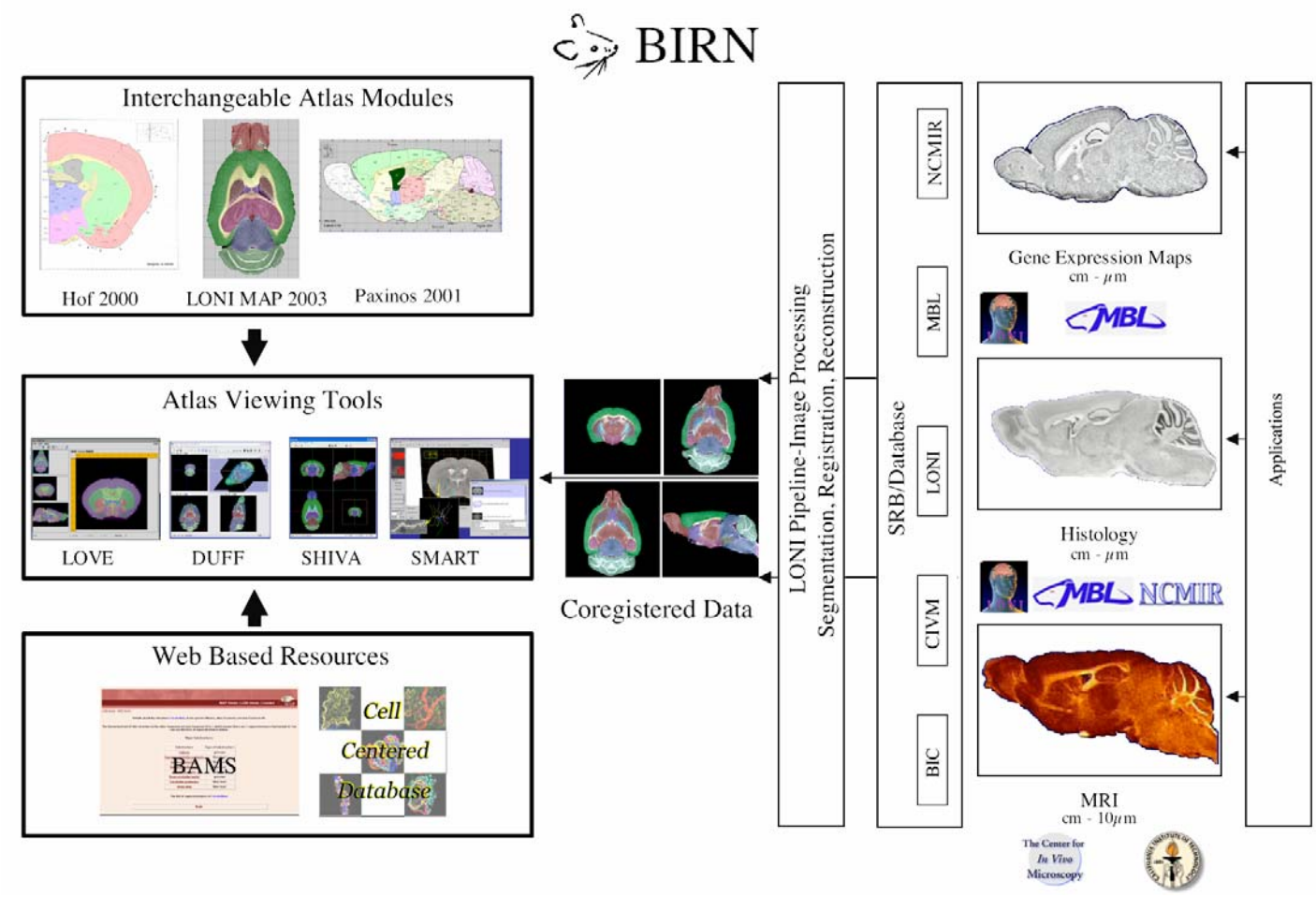
EXECUTIVE SUMMARY

This BIRN renewal application describes a research and development effort that will result in a distributed adaptive database and multiscale, multimodality atlases of the mouse brain. Tools to incorporate and compare data from gene expression patterns to gross morphology from multiple laboratories at scales from nanometers to centimeters will be built. This is an effort to integrate data already existing or to be generated as part of this or other NIH-wide or privately funded programs. The goals are to create an infrastructure for the mouse brain that can greatly enhance these other efforts by relating previously disparate data collections into a single system capable of quantitative visualization and linkage with previously disconnected knowledge bases.

This is not a project to create a single large database. Rather it is a project that allows a linkage between data, connecting diverse research groups and databases. It will result in a suite of tools that can adapt to unique data collections, from large population studies to incomplete sparsely sampled individual samples. It will integrate diverse data types from image volumes to HTML and relate them with well defined ontologies. It will provide the framework to accommodate new data and new information. It will be a test bed for new database, atlasing, visualization, and image processing tools for large scale computational efforts.



Executive Summary, Figure 1: BIRN Overview. An overview of the design of the BIRN project.

How would it work? Once completed, researchers will be able to directly compare their data across experiments, across laboratories, and across imaging modalities. For example, someone interested in experimental autoimmune encephalomyelitis (EAE) will be able to import their magnetic resonance imaging (MRI) data into an EAE atlas and compare it in the same space to immunohistochemistry data acquired in a collaborator's laboratory. They could compare their data with neuroanatomical changes, histological correlates of lesions detected by MRI, or gene expression in a region of interest (ROI). They could query to see what other genes are expressed in their ROI, visualize in situ hybridization images of those genes collected in a third laboratory (still in the same space), and then query to see what genetic loci are associated with susceptibility when compared to a resistant strain. And then repeat the process with the resistant strain. This proposal focuses on time-varying neurodegenerative diseases and on developing tools to not only aid in answering

these questions, but in asking more powerful ones. This project will not solve the neurobiology of disease, but will produce the infrastructure to enable us to pose and answer these and other more challenging questions.

At first sight this may seem an insurmountable task. However, the ☆BIRN builds upon considerable computational and informatics infrastructure created during the last 2.5 years of funding. We have successfully developed, deployed and tested a variety of tools to convert different data types such as high field MRI, blockface imaging, histology, *in situ* gene expression maps and light and electron microscopic volumes. We have engineered preliminary databases, coded several image viewers, and developed processing tools to prepare the data. We have built and tested several atlases. Each of the participants in the project is a well-respected expert in relevant fields with considerable experience performing the basic science that compels this project.

OVERVIEW OF THE BIRN PROJECT

The Biomedical Informatics Research Network (BIRN) was launched in 2001 with the goal of fostering large-scale collaborations in biomedical science by utilizing the capabilities of the emerging cyberinfrastructure (high speed networks, distributed high-performance computing and the necessary software and data integration capabilities). In its initial phase (2001-2004), the BIRN involved a consortium of 12 universities and 17 research groups participating in three test bed projects (Function BIRN, Morphometry BIRN and Mouse BIRN) centered around brain imaging of human neurological disease and associated animal models. The two human test beds, Function BIRN and Morphometry BIRN, work on large scale, cross-institutional imaging studies on Alzheimer's disease, depression and schizophrenia using functional and structural MRI respectively. Mouse BIRN studied animal models relevant to the study of multiple sclerosis, attention deficit disorder and Parkinson's disease through coordinated application of multiscale imaging technologies including MRI, whole brain histology and high resolution light and electron microscopy. These studies were used to drive the definition, construction, and daily use of a *federated data system.* Federation presents biological data held at geographically-separated sites to appear as a single, unified and persistent data archive. Data is securely accessed across institutional boundaries to address issues of data privacy and automatic translation of data formats. Most of the groups participating in the BIRN have traditionally conducted independent investigations on relatively small populations, using site specific software tools. The promise of the BIRN is the ability to test new hypotheses through the analysis of larger patient populations and unique multi-resolution views of animal models made possible by data sharing and the integration of site independent resources for collaborative data refinement.

An essential feature of the BIRN project is the collaboration of neuroscientists and computer scientists as equal partners. It has required that both groups be interested in and committed to learning the language of the other to enable understanding of their needs and constraints so that both can work together effectively to make progress on common goals. The variety of perspectives (biological, policy, information technology) and the high level of interaction among these groups inform every step of the design and implementation of the distributed architecture. The test bed projects present practical and immediate requirements for performing large-scale bioinformatics studies and provide a multitude of usage cases for computation and data in an inherently distributed environment. While each of the test beds shares the BIRN cyberinfrastructure, each imposes unique requirements derived from its specific scientific aims.

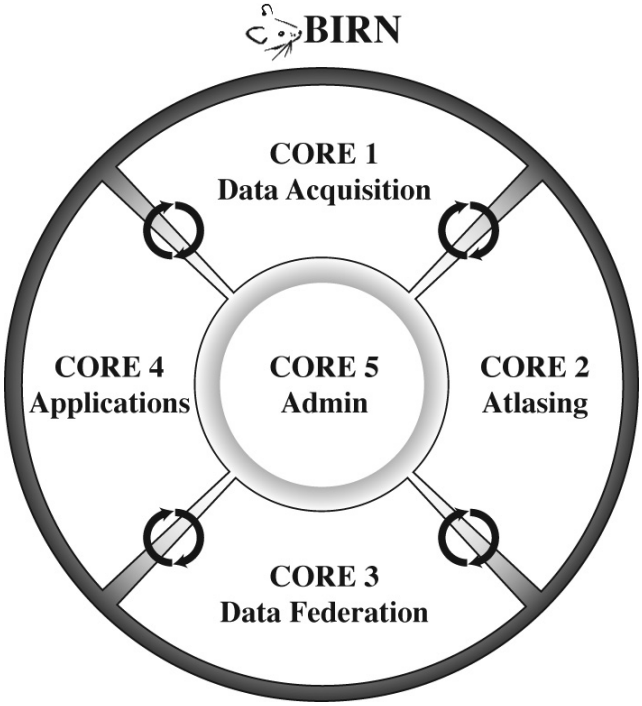
☆BIRN STRUCTURE

☆BIRN will integrate the activities of five existing laboratories – the Laboratory of Neuro Imaging (LONI) at the University of California at Los Angeles (UCLA) in collaboration with the MRI Center in the Biological Imaging Center (BIC) at the California Institute of Technology (CIT), the Center for In Vivo Microscopy (CIVM) at Duke University, the Mouse Brain Library (MBL) at the University of Tennessee Health Sciences Center (UTHSC), the National Center for Microscopy and Imaging Research (NCMIR) and the National Biomedical Computation Resource (NBCR), both at the University of California at San Diego (UCSD).

This competitive renewal also has a change in P.I. Al Johnson at Duke was PI of the previous effort. As the scientific focus has matured, it was felt that Arthur Toga should assume the PI-ship to help guide the effort during its next phase. This was a group decision.

The ☆BIRN is structured as five cores with participants from all sites, each contributing their expertise and experience to several of the cores. The cores are titled: Data Acquisition, Atlasing, Data Federation, Applications, and Administration (Figure 2).

The Data Acquisition Core will acquire data for the entire project, encompassing imaging modalities from the whole brain scale to the supramolecular and microarray-based expression map studies. The Atlasing Core will enable the processing of imaging data, reconstruction and registration of it, using atlasing techniques to integrate the various data collected into multimodal digital atlases. The Data Federation Core will organize, manage, and archive all the data collected and develop mechanisms for interaction between databases. The Applications Core contains the neurodegenerative disease test beds, the research projects that drive the development of an infrastructure designed to enhance multisite collaborative studies on animal models of disease. And finally, the Administration Core will manage communication between and within cores, to the BIRN Central Coordinating site (BIRN-CC), and to the scientific community at large. As work on these prototype projects progresses, we will evaluate additional collaborative projects in animal models of neurological disease, which could be enhanced through collaboration between laboratories. The infrastructure developed in animal models of neurological disease will lay the foundation for future extensions.



Executive Summary, Figure 2: Core Overview. A schematic describing the structure of BIRN.

CORE 1: DATA ACQUISITION

This proposal hinges around the integration and archiving of large and disparate imaging data sets. The Data Acquisition Core will develop and acquire multimodal-multiscale imaging data sets and microarray-based expression map data. The image data will be generated by magnetic resonance microscopy, block face imaging, confocal laser scanning microscopy, and electron tomography, ranging in scale from the whole brain to cellular, subcellular, and supramolecular elements. We will utilize the imaging techniques to visualize morphological differences in the brains of mouse model systems. The data acquisition will focus on the collection of images and expression map data from the exemplar neurodegenerative diseases described in Core 4: Applications. This will require the adaptation of imaging strategies so that they are specifically tailored to fit the disease models being studied.

CORE 2: ATLASING

The task of integrating the enormous amounts of data that this proposal will generate is far from trivial. Integrating disparate forms of data eludes traditional models of visualization and data collection. We propose to create multimodal digital atlases as frameworks to facilitate the organization and display of BIRN data sets. This will also entail the development and integration of various interface and image processing tools for the visualization and automated analysis of the data collected. The tools developed will be tightly integrated with the Core 3: Data Federation to facilitate the rapid and facile enrollment and retrieval of imaging data.

CORE 3: DATA FEDERATION

Diverse approaches to studying the mouse brain produce equally diverse data types, differing in modality, scale and domain. The challenge is to connect these data such that each contributes toward a more comprehensive whole. Multi-scale linkage of these data will allow the investigators a comprehensive basis for interpretation of the signals from the whole brain images relative to the tissue and cellular alterations occurring in these animal models of three distinct forms of neurodegenerative diseases.

The goal of the Data Federation Core is to build and disseminate data management tools that ease the data storage, sharing, discovery and visualization processes. To achieve this goal, the Data Federation Core will continue development of multi-scale databases and data integration along existing lines extending local database schemas, as needed, to support

additional data collections and domains. We will explore and develop alternative integration approaches based upon peer-to-peer and shared file systems. We will expand the BIRN knowledgebase by integrating additional data sources and incorporating additional ontologies, with a concentration on sources of gene expression information. These additions will help provide valuable insights into the molecular biology underlying degenerative brain disease. And finally, we will develop tools integrating data management, presentation, visualization and processing such that the investigator can query, process, analyze and visualize data across all data sources.

CORE 4: APPLICATIONS

In order to guide infrastructure development, insure that this infrastructure is tested using real world problems and to demonstrate the utility of the resulting tools, we will center the neuroscience of this project around degenerative brain disease. We chose this focus for several reasons. First, there are degenerative diseases such as Alzheimer's disease (AD), multiple sclerosis (MS), and Parkinson's disease (PD) that result in characteristic morphological changes that can be detected in vivo and histologically. There are effects in gray and white matter that can be measured, catalogued, and visualized. Second, these diseases have mouse models that could greatly benefit from the integrative approach proposed here. Genetics studies that rely on genetically engineered mouse models (transgenic lines, knockouts and knockin mice) are now maturing beyong simple qualitative assessment and now demand quantitative comparisons from the molecular to whole brain level and often on several strain backgrounds. Third, AD is the focus of MorphBIRN and a synergy afforded by comparisons between these two BIRN efforts will be amplified with a common disease test bed. The Experimental Autoimmune Encephalomyelitis (EAE) model of MS and the alpha-synuclein knock-out model of PD examined during the previous funding cycle demonstrated the feasibility and potential of this model to utilize the tools and infrastructure proposed here. Lastly, among the neurological diseases, these three are among the most debilitating disorders occuring relatively frequently in populations, thus justifying their selection as testbeds.

CORE 5: ADMINISTRATION

Successful implementation of the BIRN project requires close cooperation among dozens of neuroscientists, computer scientists, database engineers, grid computing engineers and others. In order to more efficiently organize and integrate these efforts we have divided the project into a number of inter-dependent cores. The Administration Core will be responsible for coordinating the efforts of the other cores and sponsoring educational outreach programs. It will be responsible for addressing the many and complex logistical, administrative, fiscal and coordination issues associated with the project and its new mandates.

The Administration Core will track the progress of software development as well as the validation and testing of all aspects of this program. Broad dissemination of the newly developed infrastructure, software and data resources are of the utmost importance. Every aspect of the design of this proposal and the intent of the investigators on this project is oriented towards the dual goals of achieving excellence in our work and having it support the widest possible community of users. To this end the Administration Core will aggressively pursue external entities to encourage them to interact with us and utilize the tools that result from our efforts.

SUMMARY

Our goal, in this next phase of the project, is to produce integrated frameworks for the collection, analysis, management, and visualization of neuroscientific data. We propose to produce a distributed adaptive database and multiscale, multimodality digital atlases as a method of addressing these issues. We will use neurodegenerative diseases in the Applications Core as test beds to spur the development ofdata acquisition, atlasing, and data management tools. Novel imaging technologies will be developed by the Data Acquisition Core to visualize the morphological changes caused by these diseases. The Atlasing Core will develop image processing and visualization tools for the automated processing and analysis of large data sets. The Data Federation Core will develop tools to integrate the disparate forms of neuroscientific data and allow for their facile organization and query. And all these tools will be integrated with one another, visualization applications with databases, databases with image processing tools, image processing tools to data collection tools, producing a suite of tools not only available to researchers at large, but actively disseminated by the actions of the Administration Core. In this way we endeavor to develop an inclusive model for neuroscience, achieving a truly cooperative and interoperative exchange of data and tools.