The INCF Digital Atlasing Program

Report on Digital Atlasing Standards in the Rodent Brain

incf

INCF Working Group on Digital Atlasing

Lead

Michael Hawrylycz, Allen Institute for Brain Science, Seattle, USA

Project Coordinator

Jyl Boline, Informed Minds, Hollywood FL, USA

Members

Albert Burger, MRC, Human Genetics Unit, Edinburgh, United Kingdom

Tsutomu Hashikawa, RIKEN Brain Science Institute, Saitama, Japan

G. Allan Johnson, Duke, Center for In Vivo Microscopy, Durham, USA

Maryann Martone, National Center for Microscopy and Imaging Research (NCMIR), University of California, San Diego, USA

Lydia Ng, Informatics Group, Allen Institute for Brain Science, Seattle, USA

Jonathan Nissanov, Drexel University School of Medicine, Philadelphia, USA

Luis Puelles, Faculty of Medicine, University of Murcia, Spain

Seth Ruffins, Laboratory of Neuro Imaging (LONI), University of California, Los Angeles, USA

Fons Verbeek, Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands

Ilya Zaslavsky, UCSD Supercomputer Center, San Diego, USA

Supported by the INCF Central Fund and the Swedish Foundation for Strategic Research.

Contents

1 Executive Summary

The goal of the *INCF Digital Atlasing Program* is to provide the vision and direction necessary to make the rapidly growing collection of multidimensional data of the rodent brain (images, gene expression, etc.) widely accessible and usable to the international research community. This Digital Brain Atlasing Standards Task Force was formed in May 2008 to investigate the state of rodent brain digital atlasing, and formulate standards, guidelines, and policy recommendations.

Our first objective has been the preparation of a detailed document that includes the vision and specific description of an infrastructure, systems and methods capable of serving the scientific goals of the community, as well as practical issues for achieving the goals. This report builds on the 1st INCF Workshop on Mouse and Rat Brain Digital Atlasing Systems (Boline et al., 2007, *Nature Preceedings*, doi:10.1038/npre.2007.1046.1) and includes a more detailed analysis of both the current state and desired state of digital atlasing along with specific recommendations for achieving these goals.

1.1 Overview of Digital Atlasing and its Relationship to Neuroscience

Digital brain atlases are useful as references, analysis tools, and as data integration frameworks for applications in neuroscience. Atlases and supporting tools are crucial resources in promoting data sharing. With the advance of informatics in the life sciences, neuroscientists have advocated the use of digital atlases as gateways to data of multiple modalities and from distributed locations. This infrastructure places digital atlases at the hub of a system that allows easy access to data and tools from multiple sources, and in addition, facilitates scientific discovery, data analysis and standardization (see Figure 1 on p. 13). The system proposed in this report would be both helpful to individual neuroscientists and a powerful tool for furthering neuroinformatics research.

Several projects are connecting infrastructure to the spectrum of neuroinformatics tools being developed, and these facilitate sharing, managing, and retrieving data of different types, scale, and even location. "With these [tools] in place, we have the ability to combine, analyze, and interpret these data in a manner not previously possible, opening the door to examine issues in new and exciting ways, potentially leading to speedier discovery of answers as well as new questions about the brain." (Boline et al., 2008)

1.1.1 Goals for this Framework

The high level goal of this project is to design and create an atlas-based data sharing hub that will allow researchers from individuals to large groups to share data from a variety of rodent based experimental modalities from distributed locations, and ultimately be able to view and perform comparative analyses of the results. However, atlas integration requires the ability to specify anatomic location which can be specified in at least three ways, including:

1. atlas coordinates
2. ontology (e.g. tissue label)
3. spatial placement rules, i.e., a defined region of interest (ROI) specified as a collection of spatial rules that sequentially narrow the ROI space, e.g. "within N microns from tissue1" and "between tissue2 and tissue3" and "adjacent to tissue4".

We recognize that different types of documents are likely to rely on different location descriptions. For example, neuroscience publications may contain verbal descriptions of signal or tissue location, which may be represented as spatial placement rules, while currently available atlas query systems may return a tissue label given a pixel location in some coordinate system, or vice versa. The ability to translate between different location description modalities is an important component of atlas interoperability, ensuring that different types of documents are integrated in comprehensive ROI descriptions. However, before different modalities can be considered we shall focus on formal metrics of the space of rodent brain. Therefore, in the architecture proposed in this document, we focus on atlas coordinates to aid in atlas integration, though we expect to also incorporate ontology in the near future. This integration will be made possible by taking advantage of ongoing efforts to create ontology infrastructure, including those by INCF (e.g., *Program on Ontologies of Neural Structures*). The combination of these two will be used for specifying location via spatial placement rules.

a. Addressing User Needs

A standardized digital atlasing system must be usable both for biologically and computationally oriented people. It must also be built in a flexible and extensible manner that will allow adoption by other communities. There are multiple scenarios for how this environment may be used based on the needs of the user, and several such use-cases are outlined in Appendix C. The most likely need, as well as potentially beneficial result of this framework, will be to provide researchers with access to data resources and tools that complement their own research (see Representative Use Cases in this section of the report).

Throughout all aspects of development, community feedback and participation is crucial both for usability as well as adoption. Interactions with the community will be through publications, presentations, workshops, and projects that highlight the advantages of the system and that facilitate feedback on areas that need improvement and development.

b. Requirements of the System

In summary, the requirements for this system are that atlases should act as the map to which data of both multiple types and sources may be linked both for processing, upload, analysis, data retrieval and sharing (Figure 1). All components of this system must be freely available to anyone, easily accessible, and straightforward to use. The system will require both spatial and semantic normalization which will require the creation of best practices, standards, and services. Additionally, the system must be created in a manner that is both extensible as well as flexible enough to be adopted by others. A user should be able to use a variety of interfaces that fit their needs, thus tools must be created that adhere to the standards of the system. Finally, users should be able to easily publish their findings and collaborate with others.

Of these standards, we stress the importance of both spatial normalization and standardized terminologies and ontologies, as these are the basic building blocks on which atlas integration for data sharing and visualization revolve. As the focus of this group is digital atlasing, we will concentrate on spatial normalization for this report, but reiterate that semantic normalization is equally important to this and other data sharing efforts. Thus, we want to ensure that the infrastructure developed in this area by the *INCF Program on Ontologies of Neural Structure* (PONS) group also fits the needs of the digital atlasing community.

c. Practical Considerations

As this vision will require a significant community effort, we have broken down our recommendations into immediate and long-term goals. The immediate recommendations create a normalized space and a method for interacting with that representation, while our full vision is outlined in the section of the Report on long-term goals and specifications of *INCF Digital Atlasing Infrastructure (INCF-DAI)*.

While the rodent imaging community is small in comparison to the human, digital atlasing in the rodent still represents a large body of work. We recommend that the initial focus should remain on the adult C57BL/6J mouse given the wealth of data available for that model organism and its significance in genetic studies. After development of a prototype, the system should be expanded to other strains, developmental stages, and species over time.

1.2 Immediate Recommendations

There is prior work in connecting some of the key resources used in the community; however, one of the unique features of the present effort is the recommendation of an overarching architecture that will provide for systematic interchange of data sources and the ability to add new entities in a reasonably straightforward manner.

Our first principal recommendation is to create an atlas framework that would provide interconnectivity between several existing key neuroinformatics resources, and in principle would allow any group to add new resources. The central paradigm is based on a canonical atlas space, called *Waxholm Space*[1] (WHS) enabling standardization in mapping and data access. In this section of the Report, we present a plan for how to create this space and render resources compatible with it. In addition, we also present several related recommendations, such as the dissemination of certain best practices, and a focused task force to create, implement and test the first instance of this framework.

1.2.1 Standardization in Atlas Mapping: Waxholm Space (WHS)

While the idea of using a reference space for the canonical mapping and registration of anatomic based data is not a new idea, there is currently little standardization between existing atlasing systems and workflows. A natural starting point for improvement would be for atlasing groups to map their atlases into a coordinate based standard space, targeting first those with large associated data sources. This new canonical space (WHS) provides a common spatial standard that allows translation between different digital atlasing efforts (see Figure 2) and the integration of existing key neuroinformatics resources. Standardization will allow for future efforts to do the same. The core principles of the WHS standard are that:

- Several best practices for digital atlasing exist and can be realized through standardization in mapping and data access.
- Rigorous construction of a default architecture supporting data mapping (WHS) enables interoperability and derives maximum utility from existing data repositories.
- As central data repositories and atlases are linked through a common standard there is a greater opportunity for enhancing the research process, as well as encouraging new data providers to adopt the standard.

In this report, we first propose a definition for this standard canonical atlas space; we then propose a plan for "building" WHS, followed by the registration of key reference atlases into this space. Simply put, WHS is a continuous Cartesian coordinate system (with the origin as illustrated in Figure 3), and with a virtual stereotaxic orientation in order to remain consistent with paper atlases. WHS is essentially the equivalent of the Talairach space for humans and will enable comparable standardization and comparison between rodent specimens. Once a dataset is in Waxholm Space, it is trivial to change the orientation and origin. WHS standardization in the C57Bl/6J mouse has the potential for greater relative scientific benefit than the corresponding identification of Talairach space in the human. This derives from the comparatively low variability in phenotype amongst laboratory model organisms.

The initial WHS standard construction will be based on high quality data sets to insure the greatest downstream accuracy. We propose collecting high-resolution MRI data with multiple contrast mechanisms and associated Nissl volumes specifically for this purpose. This is because MR at microscopic resolution can provide a consistent, undistorted 3D reference frame to which Nissl and other data types can be mapped. More details about this dataset and procedure are given in the Constructing Waxholm Space section of the Report. Using the MRI volumes, an average brain and probabilistic atlas in WHS will be created, with structural surfaces consistent with the high resolution MRI. The associated Nissl volumes can inherit Waxholm Space coordinates through alignment to the probabilistic MRI or the MRI volume that was collected from the same animal, or it can be brought into WHS through a similar process as the MRI volumes.

Once the standardized WHS is populated, key community datasets and reference atlases can also be brought into this space. The first set of resources that we recommend are those that have high utility data tied to them (specific atlases are listed in this section of the Report with more detail in Appendix F). These atlases, along with their registration transformations will be made available to external groups in order to make their data and associated services accessible to others.

[1]Waxholm, Sweden, in honor of the INCF meeting location where this architecture was proposed.

1.2.2 WHS Task Force

The scale of the proposed effort involves substantial additional research, architecture, and design. Much of the WHS architecture will require an additional technical task force specifically concerned with the design and implementation decisions. This will insure that WHS is created and populated with reference atlases that access data sources. This new working group will focus on refining and building WHS and create a standard application programming interface (API) for passing MRI and Nissl data from the C57BL/6J adult mouse into it. This group will also be responsible for bringing targeted reference atlases into WHS. In addition, it will set up appropriate standards and guidelines to make it easier for others to both bring their data into WHS and to access data once it is WHS compliant.

A subgroup of the Digital Atlasing Task Force will also participate in the WHS Task Force. However, this group should include other individuals with a great deal of practical experience with registration as well as tool-builders that would be interested in making use of the spatial transformation information.

1.2.3 Immediate Goals for the Digital Atlasing Task Force

Immediate actions by this task force include developing a set of best-practices for making it easier to move different types of data into WHS followed by continuing work on refining and prioritizing the components of the longer-term infrastructure goals.

1.3 Longer-Term Goals for Interoperable Digital Atlasing Infrastructure

1.3.1 Overview of INCF Digital Atlasing Infrastructure (INCF-DAI)

While the immediate recommendations above advance digital atlas interoperability, a larger system needs to be put into place to meet the full vision put forward by this group. We have identified such an infrastructure based on work in other areas and researchers' needs, and it is outlined in this section of the Report. We have also begun analyzing current resources to see how they might fit into such an infrastructure, and to identify where key items are missing. This is an ongoing process and we hope to identify many more specific recommendations in this area in the near future.

a. Requirements

In order to facilitate the building of interoperable infrastructure, underlying standards must exist; however, these standards shouldn't constrain development, instead there should also be a system built that track data about the various atlases, and convert the information in a tool-accessible manner.

In order to create this infrastructure, an architecture document and a forum for discussion and improvement of this architecture must be created. These will allow members of the community to contribute their ideas and build specific components of the system.

b. Architecture Vision

INCF *D*igital *A*tlasing *I*nfrastructure (INCF-DAI) is envisioned as a collection of distributed services that support the publication, discovery, and invocation of heterogeneous atlas resources (Figure 4, page 19). INCF-DAI will reference remote and autonomously supported and updated resources, and host or mirror some resources that are critical for operation of this infrastructure. These resources will include atlasing datasets of different types (see Figure 5, page 20), compute resources, applications, and workflows. INCF-DAI will establish common access mechanisms for resources of each type, and provide and govern respective standard API development, to ensure syntactic interoperability within the system, and support development of various specialized atlasing clients over the common services framework.

The atlasing integration framework will further rely on common space conventions (WHS) and semantic conventions, each of which will be supported by respective integration servers. By relying on a uniform set of access APIs, INCF-DAI will be scalable in terms of data and compute resources that can be made available within the system. This architectural approach will support and encourage easy addition of INCF-DAI-compatible resources developed by researchers in different organizations, making it a community infrastructure.

The construction of the technical infrastructure for the WHS standard must facilitate interoperability between atlases and data sharing, yet allow for reasonable extension and flexibility. Underlying standards for independent atlas builders should exist as recommendations for optimal construction practice; however, it is impractical to constrain developers by attempting to enforce standards. Instead, the architecture of the standard should be such that it supports mapping between different world views in a data accessible manner.

From an architectural and technical perspective the construction of the WHS standard can be accomplished independently of the INCF-DAI, at least in the initial stages. The main points for the WHS are to establish a standard coordinate system and to identify the central data modalities that can be mapped and are sufficiently rich enough to encompass modern digital atlasing applications.

c. INCF-DAI Specifications vs. Implementation

The architecture vision in this report is the first step towards creating a more detailed specifications and architecture document. The development of these will aid in identifying specific areas where this program, INCF, and the rest of the community may most effectively play a role in filling these key resources.

The primary role of this group is to develop these specifications for INCF-DAI rather than to actually implement the recommended infrastructure. Implementation of this infrastructure is beyond the scope of the INCF-DAI and requires a larger community effort that is funded by the member country nodes of INCF.

1.3.2 Towards the INCF-DAI Specification

a. Prototype

As discussed earlier, atlas interoperability is to be supported by atlas coordinates, ontology (e.g. tissue label), and spatial placement rules. WHS addresses the first of these methods, but the latter two modes of integration are critical where brain images vary greatly from this standard. They will also play an important role when dealing with the temporal aspects of brain development and the integration of related data sets. In addition to a precise definition of the conceptual Waxholm Space and its realization as a standard canonical atlas, we also propose a project to both test and begin the development of referencing atlas location by tissue label and by spatial placement rules. This project will focus on using WHS to integrate the Edinburgh Mouse Atlas with other developmental atlas data, specifically the Allen Brain Atlas, but time permitting, other atlas data.

Again we note the goal is not to implement the complete INCF-DAI, as this would be substantially beyond the scope of this work, but to ground the development of the specifications of DAI components dealing with developmental data and semantic atlas integration in concrete examples (see Figure 6).

b. Community Engagement

As in all aspects of this program, it is crucial that the interested parties in the scientific community are involved in multiple steps of this process via both in-person workshops and web tools for on-line feedback and discussion.

1.3.3 Towards INCF-DAI Implementation

Again, the goals of this group are to specify rather than implement INCF-DAI. However, it is both desirable and likely that at least some prototype pieces of this infrastructure will be developed during its specification. Any usable components that are developed will be shared with the community through INCF. As with any complex infrastructure, we expect the implementation of INCF-DAI to be iterative, due to new events and issues that arise during development.

1.3.4 Connecting Key Resources to WHS through INCF-DAI

The connection of key resources to WHS via INCF-DAI serves two purposes. First of all, it links valuable resources and information to the environment that has been outlined in this document. In addition, the practical goal of linking these through INCF-DAI aids in developing the specifications and standards of the infrastructure.

The goal of complete interoperability with all existing resources is of course not feasible. However, an outcome of this working group is the identification of key core resources that are:

- Identified as central neuroinformatics atlases or data repositories in rodent digital atlasing
- Candidates for basic interfaces to the central WHS standard
- Operate interchangeably with high interconnectivity
- Serve as basic models for connecting new resources

We have identified some key resources that should be considered for linkage to WHS and the INCF-DAI. A list of these is presented in this section with more detailed information in Appendix F. Initial linkage of these resources should fall to the WHS Task Force.

1.4 Long Term Infrastructure Implementation Recommendations

We believe that the present group can continue to lead the main design and architectural decisions central to the establishment of the WHS standard and INCF-DAI. However, in order to tackle tasks specific to a community, experts from those communities will be required. In addition, certain tasks will require technical people to help develop, implement, and test the practical incorporation of standards and APIs into usable software.

We have proposed two projects to quickly make headway on creating some key components of this infrastructure and to aid us in further defining INCF-DAI. These in conjunction with ongoing work by this task force will allow us to create an architecture document for infrastructure components. Once this is created, it will be much easier for community members to focus on contributing to specific components of the INCF-DAI. It is also more appropriate to form specific ad-hoc committees to focus on specific components of the infrastructure.

1.5 Why the INCF?

The INCF is uniquely positioned to facilitate a multi-national and multi-resource collaborative activity of this nature. It is clear, however, that the infrastructure pieces should be built by the informatics members of the scientific community along with the individual resource stakeholders with guidance and support in crucial areas from INCF. Ultimately, as the system develops and later matures there will be more opportunities for the integration of new resources. The INCF may act as a clearinghouse for these technologies and datasets to enable the system to maintain an active and living character.

2 Report

2.1 Process

The INCF working group investigation leading to this report and its recommendations was conducted via online meetings and exchanges starting in May 2008, and with a digital atlasing task force two-day meeting in Waxholm, Sweden, held September 5-6, 2008. The main ideas of the architecture and system were solidified at this meeting. Follow-up online meetings were held to complete this report, though we continue to develop infrastructure and plans. A basic outline of this task force's process is as follows:

- Identify the goals for this task force as well as digital atlasing in general
- Identify atlasing use cases, typical workflows, and research scenarios for different research needs
- Apply the use cases and workflows as a guideline to identify key tools and resources for this effort; these include:
  - existing data sources and canonical datasets
  - protocols (and tools for pre-processing)
  - spatial registration
  - database upload
  - annotation and markup
  - query and access
  - analysis
  - integration resources
- Generate recommendations for a canonical digital atlasing space, including how to create, populate, and use this space
- Identify a vision for INCF Digital Atlasing Infrastructure (INCF-DAI) along with guidelines for achieving this infrastructure
- Identify initial projects to aid in creating and testing this space (INCF-DAI)
- Generate several "best practices" documents with additional standards
- Identify key areas where specialized task forces are needed to create standards, and implement specific test cases to further define an INCF-DAI architecture document.

2.2 Overview of Digital Atlasing and its Relationship to Neuroscience

The organization and understanding of the massive data influx occurring in contemporary life requires powerful mapping tools to access and process its significance. In nearly every aspect of daily life we are inundated with data from information on directions, restaurants, traffic, houses for sale, to financial and medical records. Atlasing and visualization is becoming more of a standard and expected option for managing this information. With barely a second thought, we use these maps to examine information from different tools to give us a picture of an area and to analyze information in a manner only recently made possible. In addition, sites such as Google have provided simple resources and an infrastructure that allow developers to easily create their own new tools. This combination greatly enhances a knowledgeable user's access to data which helps them make informed decisions.

Atlases have a long history in the life sciences from the drawings of the ancients to the first modern comprehensive atlas of the human in the work of Vesalius. With the advent of large scale digital processing capabilities neuroscientists have argued for a comprehensive digital atlas of the brain and much work has proceeded with mouse digital atlases serving as the framework used to allow a user to traverse the brain and information linked to it (Bjaalie, 2002, Toga 2002, Baldock et al., 2003, MacKenzie-Graham et al. 2003, Martone et al. 2004, Boline et al. 2007, Boline et al. 2008). Brain atlases may be used as an intuitive portal interface to knowledge as the visual and mental model of an atlas is very appealing to users.

Because digital brain atlases are useful as models, references, and analytical tools there are several digital rodent atlases and related tools that have been developed by the neuroscience community for the various purposes:

- teaching tools
- to convey methodology (e.g., developmental, functional) for parcellating the brain
- to indicate locations of relevance for function or structure, for example in clinical or research applications
- to compare phenotype (and sometimes associated genotype or functional) differences between
  - different time points or conditions
  - two subjects
  - a group of subjects
  - to other species

More recently, efforts have focused on using digital atlases and associated tools as a scaffold for data sharing informatics hubs. Most of these efforts use this infrastructure to share data from their own resource. This infrastructure must be expanded such that from the user's point of view, they can easily use this atlasing infrastructure to share their own data, find and analyze data from others and share their findings regardless of where they or the data are located (Figure 1). These are drivers for why advances are being made in the use of digital atlases as a framework for localizing data, especially gene expression data. Such a system would be very useful for sharing, organizing, and analyzing data, but it has not yet been created for the average science user. Multiple reasons exist for why this is not yet available for the general scientific community, including personal research objectives, lack of funds, and organization.

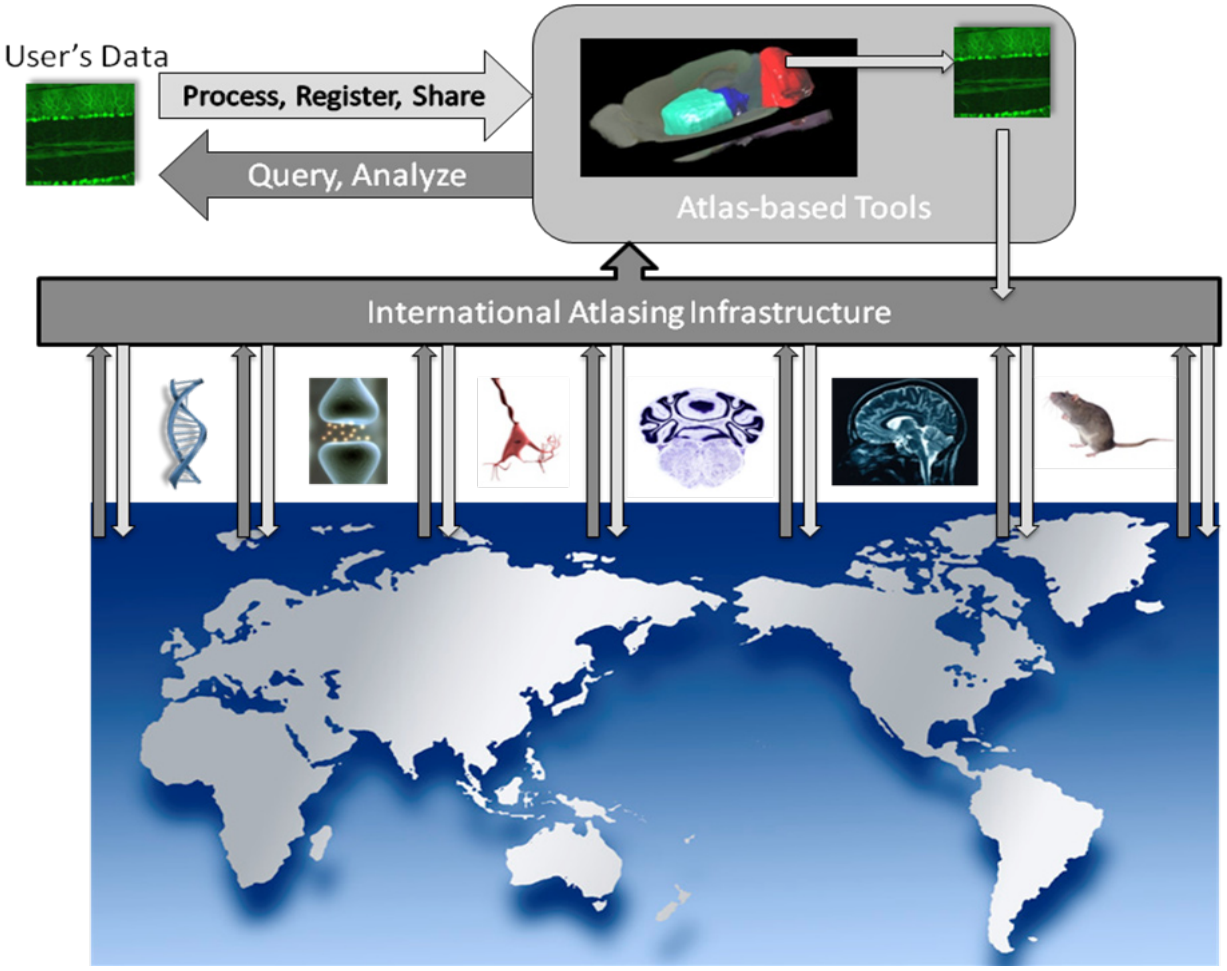


Figure 1. *Using atlas-based tools, researchers would be able to share their data and tie it to both semantic and spatial information as well as find and examine data of multiple types related to their question, from anywhere in the world.*

2.2.1 Goals for this Framework

The high level goal of this project is to design and create a data sharing hub that will allow researchers (from individuals to large groups) to share primary data from a variety of rodent based experimental modalities from distributed locations, and to be able to view and run comparative analyses of the results. The aim is to provide support for the building of a consistent network in brain research for dissemination of existing knowledge, integration of existing knowledge generation and discovery of new knowledge, both on the semantic and spatio-temporal level. In order to accomplish this ambitious goal, it is necessary to integrate a mass of huge and complicated information accumulated in a variety of research fields of neuroscience. This must be done in a structured, useful, and effective way, so that communication among researchers may be achieved easier and faster, and idea generation and exchange from informatics data may also be readily tested and advanced.

Atlas integration requires the ability to specify location, which can be specified in at least three ways:

- by atlas coordinates
  - as a fixed coordinate location relative to anatomic landmarks
  - as a region of interest (ROI; defined as a closed polygon, or a set of closed polygons over a set of reference plates or a convex shape)
  - in this case, translation may be needed between native spatial representation and other atlas coordinates
- by tissue label
  - whether the label comes from an INCF or standardized vocabulary
  - in this case, translation may be needed between the standardized label into a native vocabulary
- by spatial placement rules
  - a defined ROI specified as a collection of spatial rules that sequentially narrow the ROI space, e.g., "within N microns of tissue1" and "between tissue2 and tissue3" and "adjacent to tissue4"
  - in this case, tools must interpret tissue structures, compute areas defined by each rule, and then derive a ROI from a combiation of rules

All three of these methods may be used independently or in conjunction with each other. While the ideal system would take advantage of all three methods, we will primarily focus on atlas coordinates in this document. Most currently avail-

able atlas resources do not use spatial placement rules as a method for specifying location so we believe this is an area of focus for the future. Using tissue labels relies heavily on standard terminologies and ontologies. This group recognizes the importance of ontologies to infrastructure; however, we expect to work with the INCF Program on Ontologies of Neural Structures (PONS) to ensure the infrastructure they create can also be used to aid atlas integration. Therefore, we will only discuss the need for terminologies at a fairly high level in this document.

These latter two modes of integration are critical when variations in brain images are too substantial to allow their registration to a standard atlas coordinate space. They will play an important role when dealing with the temporal aspects of brain development and the integration of related data sets.

This task force proposes two projects to advance all three methods of atlas integration. These projects will also help pinpoint the exact infrastructure needs and will be prototypes of the infrastructure and integration of neuroscience data.

a. Addressing User Needs

It is of utmost importance that this system be built so that both biological researchers as well as computer programmers will actually use it. Thus all components of the system must be built for ease-of-use. In addition, INCF-related groups developing infrastructure must secure community participation and buy-in from the beginning. This will be done through a series of presentations, workshops, and target projects demonstrating the benefits of the information integration environment and standardization of data discovery, access, and publication protocols.

While our current focus is on developing an environment for rodent digital atlasing, the long-term goal is to create a system that may be adopted by other communities. This requires that the INCF digital atlasing infrastructure be built in a flexible and extensible manner which may be adapted to fit the needs of different communities. Thus, a system such as this may be useful for a multitude of purposes depending on both the background and goals of who is using it. Moreover, the manner in which one interacts with such a system also depends on their background and requirements. Several such use case scenarios are listed in Appendix C, and a likely common use case is outlined below.

i. Representative Use Case

User Profile:

I am a geneticist with basic knowledge of mammalian brain anatomy. I am looking at the impact of a gene on seizures in young children. I have limited experience with computers; I use my desktop computer at work for data analysis, and DVDs for storage.

Usage Scenario:

I have a series of sections from groups of control and knock-out animals. The experimental and control animals are sacrificed for gross anatomical analysis and cell morphology, followed by gene expression analysis. The phenotype in the experimental animals is broader than expected and encompasses brain regions with which I am not very familiar. The control littermates' brains are clearly normal, but I need to identify the affected regions in my experimental animals. In order to get a better understanding of how and what areas are affected, I would like to load images from my experimental brains to an atlas workspace and compare with the labeled atlas model.

With an atlasing tool that allows a comparison view, I identify the affected regions on the atlas and am able to compute area and volume changes in the areas of interest. This allows me to quantify my phenotype data and compare it to my behavioral information.

In addition, this tool gives me the ability to query the atlas for other related data and genes expressed in these regions of interest. With this knowledge, I select region-specific probes to perform in situ hybridizations, quantitative PCR, and microarray studies of the knock-out animals.

What the User Gains:

An atlas tool which links anatomy, gene expression data, literature and other spatially and semantically mapped data will give a user insight into what may be causing gene-related differences in phenotype and relate it to behavior. In addition, it provides clues for the next logical set of experiments. The visual nature of atlases also provides a mechanism to avoid arguments of nomenclature and to convey observations and ideas.

ii. Probable Workflow for this Use Case

1. The user starts with a 2D gene expression image that can be registered and compared to a canonical atlas
2. Atlas delineations are applied to the 2D image which gives the user a first look at areas that may be abnormal compared to the normal atlas
3. Additional reference datasets are accessible from the digital atlasing interface, which aid in comparing the experimental animal to other controls
4. More advanced analysis allows the user to identify and compute areas and volumes
5. Using the atlas, a region or location of interest can be selected in the interface
6. This spatial location is used to send a query about the genes of interest in this area
7. The query returns tissue information that intersects this spatial location along with a summary of the genes in that area
8. Additional information from other related resources may be obtained from this atlas interface that may aid in the investigation of this experiment

b. Requirements of the System

The INCF digital atlasing system should be built in a manner that will meet the needs of multiple users. In order for this to happen, we feel the system should be developed to fill the following requirements:

- This system must be more than a data repository; instead, atlases should act as the map to which data from multiple sources may be linked both for processing, upload, analysis, and data retrieval and sharing (see Figure 1 on next page)
- All components of the system must be freely available to anyone
- Any tools created must be easily accessible and easy to use:
  - o They should act as a gateway to which multiple sources are linked for processing, analysis, upload, retrieval, and sharing
  - o Ideally, a user should be able to choose which atlas interface they use
  - o Tool creators should be able to find and implement APIs and tool packages that help them adhere to standards
- Whenever possible, computationally intense operations should be handled by servers:
  - o For compatibility with the system, servers will offer their resources using standards identified by the INCF digital atlasing community
  - o Servers built for this system should allow access from outside sources (i.e., no firewall issues) or support functionality via web services
- This system must be created in a manner that is both extensible and flexible enough to be adopted by others
- Spatial normalization is required for atlas integration:
  - o A standardized atlas space should be defined
  - o Key reference atlases should be mapped into this space
  - o These spatial registration algorithms should be made available to others
  - o Methods should be in place to share spatially-linked data to these reference atlas
  - o Conversion services should be created to enable atlas data discovery and integration
  - o Information about registration methods and metrics for their performance should be supplied
  - o The spatial registration transforms used to move data into standardized atlas space should be stored for access and data mining
- Standardized terminologies and ontologies are required for atlas integration:
  - o Services should be created to enable tool and user access to INCF recognized ontologies and supported lexicons
  - o Services should be created to convert individual semantic descriptions into INCF recognized ontologies and supported lexicons
  - o Services should be created that allow a user to map any structural hierarchy onto an atlas
  - o Services should be created that can be used to track terminology structures across developmental stages
- The data types that users may contribute should span the spectrum of experimental types, from the molecular level to gross specimens. Standards should be assessed or created for each data type and incorporated into this system:
  - o Common data models and associated APIs
  - o Canonical relational and XML schemas
- Standardized formats for facilitating data exchange must be developed to access data and create interoperability between user interfaces and data resources
- Standards for information such as data collection methods, preprocessing provenance, and registration transformations need to be created to facilitate data contribution
- Suggested best practices and workflows for specific data types should be created to facilitate the contribution of new data to this system
- Any of a variety of interfaces should be available for query and access, related to with what the user is familiar:
  - o These interfaces should start to converge on standards that will allow them to cross different databases (e.g., standardized formats for data exchange)
  - o We suggest ensuring that a few targeted tools meet these standards early in the process
- Users will choose analysis tools that fit their needs, but these tools should adhere to standards that facilitate the sharing of findings and data that are created as a result of the analysis (e.g., probabilistic atlases):
  - o Ideally a user may examine and analyze any data available via this system
  - o Users should also be able to contribute analyzed data back to the system
- Users should be able to easily publish their findings and collaborate with others. Publication can be facilitated in the traditional sense, but also in online forums that include the ability to annotate data, share results and provide a platform for the scientific community to interact and debate these findings. A digital atlasing portal is an ideal platform for dissemination of digital atlasing information and tools and a place for this community to interact
- Standards and APIs should be developed and shared with the community in a manner similar to that used by W3C
- The core components of the system shall be thoroughly tested and monitored to ensure the services foundation is solid, and client applications can be built in reliance on

this infrastructure

i. Importance of Spatial Normalization

Spatial normalization is a key first step in applications where morphometric fidelity and spatial indexing are requirements. In conjunction with standards for mapping semantic information about data collection and processing, spatial normalization and mapping allows effective dissemination of new knowledge in an environment where it can be viewed or explored with easy cross referencing to existing knowledge bases. It also facilitates the ability to analyze data from different sources. As illustrated earlier, a system such as this might be able to help a researcher that wants to use an atlas as tool for reference, query, quantification, and analysis.

ii. Importance of Standardized Terminologies and Ontologies

Tissue labels are another method for specifying location and can be used to facilitate data exchange. Data labeled with semantic information can at least be grossly referenced to atlas space. For certain types of data (e.g. microarray), this is the only currently reasonable method for mapping it into a digital atlas environment for visualization. In addition, the goals of this group depends heavily on services that will easily link atlasing tools to ontologies, and that allows users to contribute terminologies linked to atlases. This is an area of great importance not just for digital atlasing, but for data sharing efforts in general. We want to ensure that the infrastructure developed in this area by the PONS group also fits the needs of the digital atlasing community and will work with them to achieve that objective.

c. Practical Considerations

While creating this type of infrastructure in a truly user-friendly manner is a significant effort, it is very clear that the neuroscience community would benefit immensely from such a system. This task force understands that a complete implementation of such a system requires the cross coordination and efforts of many groups and that such an effort may not be entirely feasible at this time for multiple reasons. Thus we have broken down our recommendations into immediate goals and long-term goals. The immediate recommendations create a normalized space and a method for interacting using it, while our full vision of digital atlasing infrastructure is outlined in our long-term goals. In addition, we propose projects to advance crucial components of this infrastructure and that will result in the delivery of usable resources. We also plan to extend this vision and analysis into other concrete recommendations and action items.

At some point, this framework must encompass different stages of development, disease states, and cross species. However, we feel that the initial focus should stay on the adult C57BL/6J mouse as a prototype, in particular given the wealth of data available for that model organism, and that after development of the prototype it should be expanded to other strains, developmental stages, and species over time.

2.3 Immediate Recommendations

2.3.1 Overview

There is prior work in connecting some of the key resources used in the community. However, one of the unique features of the present proposal is recommendation of an overarching architecture that will provide for systematic interchange of data sources and the ability for any group to add new entities in a reasonably straightforward manner. Many research efforts using anatomical image analysis have as their goals to identify biologically relevant objects present within image data, to provide a means to quantitatively analyze these objects, to compare the distribution of those objects to other features, and to properly annotate the objects so as to be able share this analysis in an integrated informatics framework.

Many of the neuroinformatics tools designed to achieve these goals have been developed in a fragmented manner with each resource providing access to unique data sets and analytic capabilities. Their integration would enrich the neuroinformatic network enabling queries and analysis across a number of resources. An example of an effort to link atlas-related resources is the Mouse Biomedical Informatics Research Network (BIRN) integration of the NeuroTerrain Atlas/NetOStat client, Smart Atlas, and MBAT, which illustrates that the functionality of these resources can be accessed via re-usable interfaces, so as to promote interoperation of tools and data.

Our first recommendation is to create an atlas framework that would allow any group to bring their resource into it, a canonical atlas space, entitled Waxholm Space (WHS) in reference to the September 5-6, 2008 meeting location chosen by the INCF in Waxholm, Sweden. Here we present a plan for the creation of this space and resources compatible with it. In addition, we recommend the dissemination of certain best practice documents, and a focused task force to create, implement, and test the first instance of this framework.

2.3.2 Standardization in Atlas Mapping: Waxholm Space (WHS)

a. Overview

One of the most central recommendations by this group is that atlasing groups (especially those with associated data) map their atlases into a coordinate based standard space called the Waxholm space (WHS). We propose a definition for this standard canonical atlas space, and propose a longer term plan for first "building" the space, followed by the registration of key reference atlases into this space. Once these atlases are in WHS their registration transformations will be made available to external groups. WHS provides a common spatial standard that allows translation between different digital atlasing efforts (see Figure 2 on next page).

In parallel, a task force should set spatial registration transformation standards, and ideally, another task force would generate common data models and APIs for the reference atlas data types. Together, these will greatly enhance the ability of external groups to access these reference atlases and their attached data and services.

We recommend that a single task force group oversee and work on bringing data into WHS, as well as the registration and validation procedures. This group should also oversee and help with bringing a few key reference atlases into WHS.

b. High Level Description

i. Defining Waxholm Space

WHS is a continuous Cartesian coordinate system and is essentially the equivalent of the Talairach space for humans. While the origin is arbitrary and readily transformable, we recommend that we place the origin at the intersection of the midsagittal plane (z), a plane tangentinal to the rostral (y) and another to the dorsal anterior commissure (x). Orientation can be either a virtual stereotaxic position (the flat-skull position where the height of lambda and bregma is equal) or alternatively, defined by three brain structures.

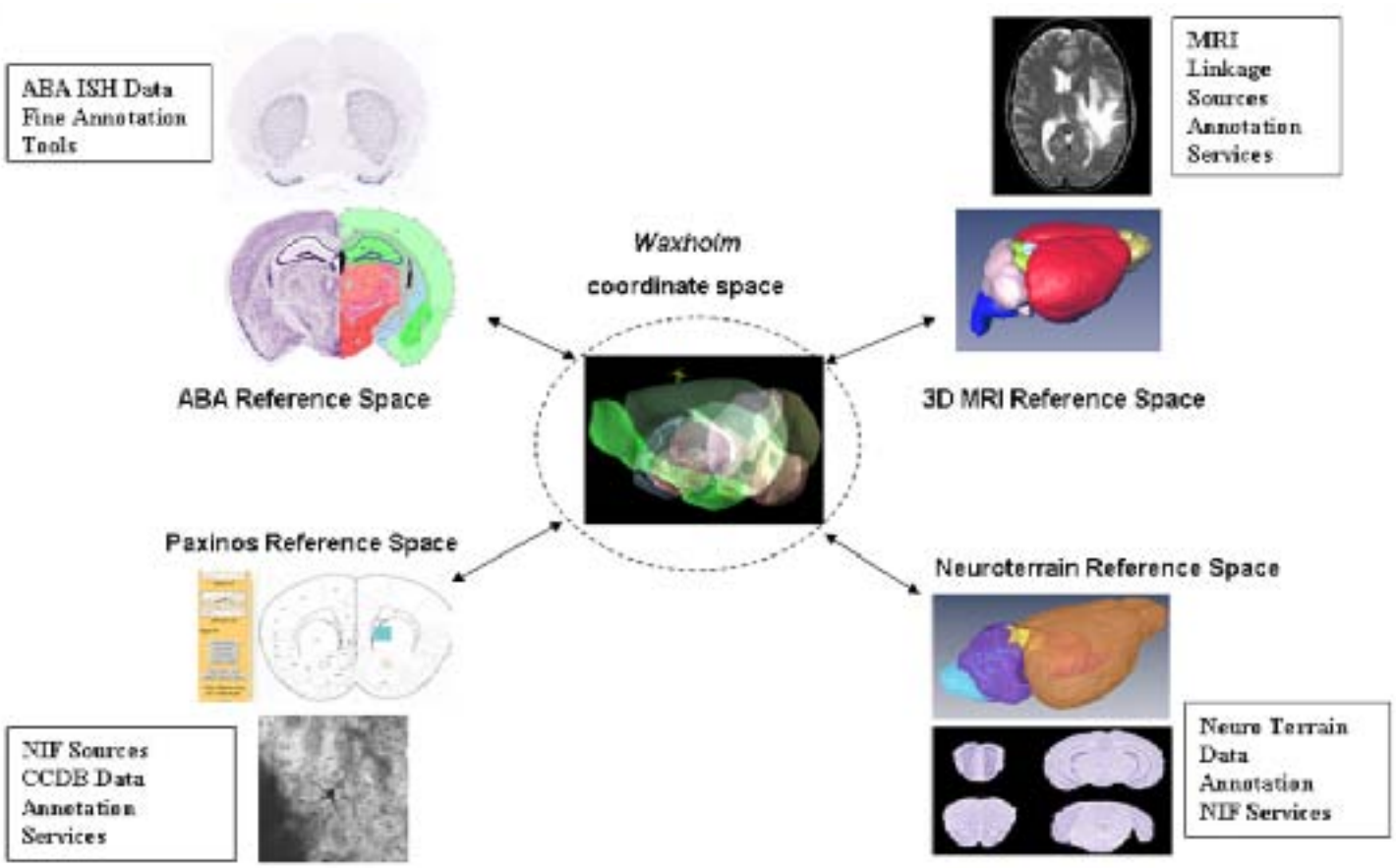
Note that neither the orientation nor the position of origin have much to do with a registration procedure and thus are arbitrary. Also, once a dataset is in Waxholm space, it is trivial to change the orientation and origin. We favor using the stereotaxic position as the origin in order to remain consistent with established mouse atlases published as hard-copies. The space should be first established for male (9-12 week) adult C57BL/6J. An example of an MR dataset in WHS is illustrated in Figure 3 on page 17.

ii. Constructing Waxholm Space

We feel the best way to proceed with building WHS is to begin by setting a standard acquisition procedure for passing data into WHS. We feel that the first standard staining procedure should be a high-resolution MR dataset with associated Nissl volumes of as many of these same subjects as possible. MR at microscopic resolution can provide a consistent, undistorted 3D reference frame which can be used as a template. Nissl is one of the most commonly collected data types and the collection process can create a great deal of distortion, thus sections for Nissl reconstruction should be collected using a minimal-distortion tape support system.

As we would like to start with the best datasets as possible, we recommend using a set of MRs collected in the skull at

Figure 2. *The canonical atlas space or Waxholm Space (WHS) acts as the hub of a centralized infrastructure connecting several key reference spaces. Reference atlases that have been mapped to this space have been "normalized" and may share their associated data and services in a manner that is understandable to external sources.*



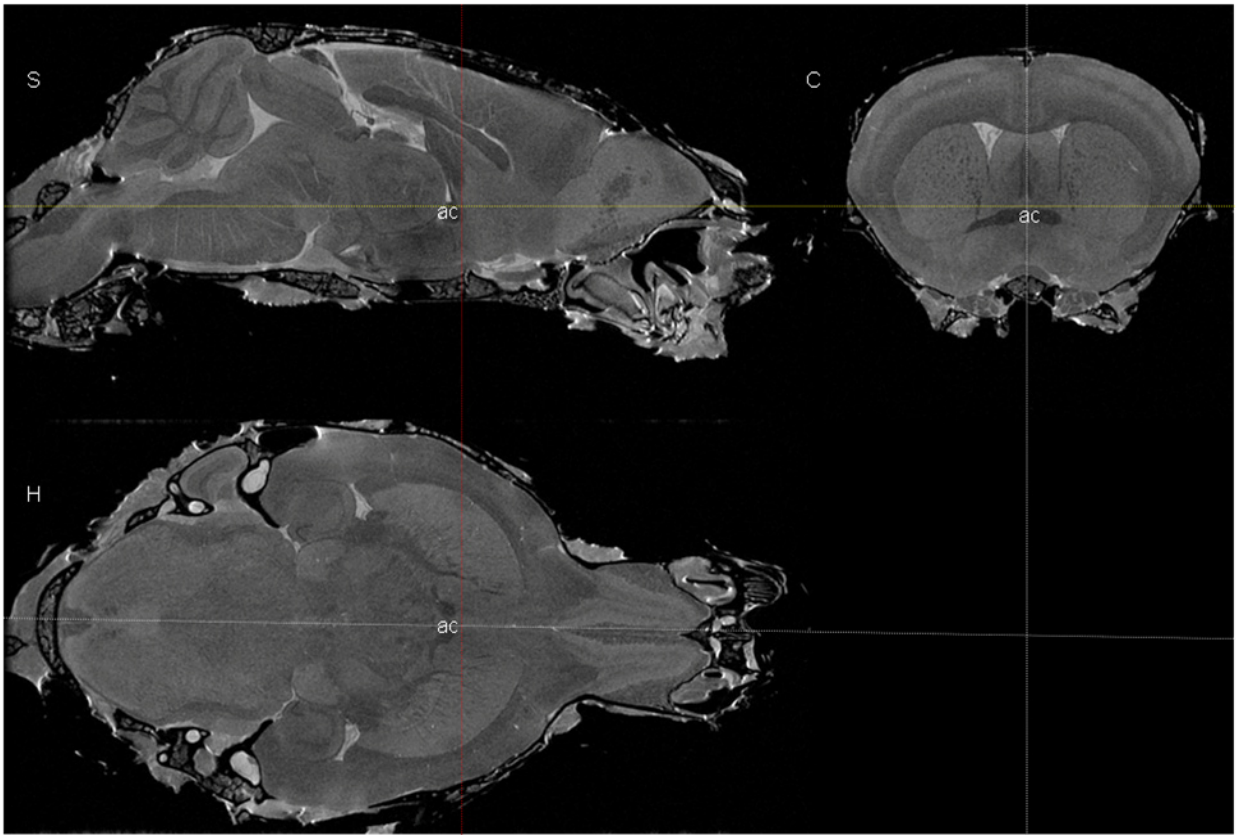


Figure 3. *The Waxholm Space origin is defined by the junction of the rostral and dorsal tangential planes of the anterior commissure (ac) with the midsagittal plane. The midsagittal plane is approximated here by S and the white dashed line, the dorsal plane roughly corresponds to H and the yellow dashed line, while the rostral plane is approximated by C and the red line. Animal orientation in the present case is somewhat offset from the specification for WS with a rostral downward and lateral shift and thus planes illustrated are approximates.*

high resolution by G. Allan Johnson's group at CIVM Duke, who has created a protocol for imaging at extremely high resolution in a manner that allows great detail of the brain. Once these specimens have been scanned they will be sent to Jonathan Nissanov at Drexel University for Nissl staining and volume reconstruction. See more details on this process in Appendix B.

Using the MR volumes, we recommend creating an average brain and probabilistic atlas in WHS along with some well-defined surfaces (e.g., cortical boundaries, external capsule, corpus callosum, internal capsule, ventricles, and the midsagittal plane). Misalignment error can be computed from the average brain of these datasets. The associated Nissl volumes can inherit Waxholm space coordinates through alignment to the appropriate MR volume.

iii. Register Key Reference Atlases into this Space

Once a probabilistic atlas is created, other groups with the help of the task force will be encouraged to also bring their data or atlas into WHS. They may either do this in a method similar to that employed for the first reference dataset, or use any of the brains or average brain(s) already in WHS as targets for registration. The first set of targeted reference atlases include those that have data tied to them, such as:

- Allen Brain Atlas
- Mouse BIRN MR atlas
- Paxinos atlas used in Smart Atlas
- Neuroterrain

Reference atlases that are in WHS should make either the transformations or indexing of coordinates available to other sources so their associated data and services may be more easily accessed by others. After these initial atlases, other reference atlases should also be brought into WHS.

We believe this process is a good place to start, and as the type of data in WHS grows, it is likely both WHS and the process will be redefined and improved.

2.3.3 WHS Task Force

We recommend that a second task force be created to ensure that WHS is created and populated with reference atlases that access data sources.

a. Mission

This group will focus on refining and building WHS and create a standard procedure for passing MR and Nissl data from the C57BL/6J adult mouse into it. This group will also be responsible for bringing targeted reference atlases into WHS. In addition, it will set up appropriate standards and guidelines

to make it easier for others to both bring their data into WHS, and to access data once it is WHS compliant. In addition, this group will share their findings with the community and solicit feedback for improvement of WHS.

There will be two main components to this task force: one experimental and the other technical.

The experimental focus:

- To generate the best possible initial datasets using standard protocols to collect data from multiple subjects. This process in and of itself will generate best practices that new groups can use as guidelines to aid them in bringing new data types into the WHS framework

The technical component consists of multiple tasks:

- Create an average brain and probabilistic atlas in WHS
- Determine if it is practical to also create a set of well-defined surfaces from this dataset that may be used as guides for new datasets
- Bring a set of reference atlases into WHS
- Develop spatial registration transformation standards, so once a reference atlas is brought into WHS, it is possible for anyone to access the registration transformations and that spatial information tied to the atlas might be located

b. People

A subgroup of the Digital Atlasing Task Force will participate in the WHS Task Force. Also, this group will need to be expanded to include a few more people with a great deal of practical experience with registration as well as tool-builders that would be interested in making use of the spatial transformation information.

2.3.4 Proposed Procedure for Constructing WHS

A total of 7 adult male C57BL/6 mice will be collected specifically for this purpose as follows:

- 3D MR microscopy images will be acquired on actively stained specimens with isotropic resolution (21.5 µm) acquired with three different imaging protocols to highlight the widest range of tissue contrast and morphology
- Volume Nissl images at 5µm x 5µm x 20 µm on the same brains reconstructed with minimal distortion and high fidelity and registered to the MR data
- A collection of 33 labels will be derived from the intact specimen's MR data, and this will be transferred to the reconstructed Nissl volumes
- These data will provide researchers with three additional methods for entering WHS:
  - Registration of 3D volume MR images with the target MR reference
  - Registration of conventional histology with the target Nissl
  - Registration of labels with the target labels

The implementation of WHS is planned for three stages and slated for completion by the end of 2009.

a. Phase I

- We will acquire the canonical dataset on a single adult male C57BL/6 mouse
  - MR data will be acquired on actively stained brains with T1, T2, and T2*-weighting at isotropic resolution (21.5 µm)
  - Nissl sections (5µm x 5µm x 20 µm) will be acquired from the same specimen
- We will register the MR and Nissl volumes and co-align them to a standardized orientation which will be defined through consensus of the Digital Atlasing Task Force
- We will provide labels for at least 33 anatomical structures and validate the ontology of these structures with accepted standards. These standards will be determined and vetted by the Digital Atlasing Task Force with the INCF Program on Ontologies of Neural Structure (PONS)
- We will place the canonical data set and the documentation defining WHS on at least one publicly accessible INCF WHS web site with suitable search, imaging, analysis, visualization and commenting tools, all of which will be freely available to the Neuroscience community
- We will define the orientation and best practices for acquisition of future data to facilitate mapping into WHS

b. Phase II

- We will expand the canonical data set by adding 6 additional data sets with identical acquisition parameters, i.e., 3 different acquisition strategies per specimen, 3 MR data, complete volumetric Nissl data, and labels
- We will register all data and create a probabilistic atlas
- We will place all data on a freely accessible INCF website

c. Phase III

- We will begin registration of representative atlases from at least three other sources to WHS starting with the Allen Brain Atlas:
- Define and begin publishing the tools and pipelines necessary for registration of MR, Nissl, and labels into WHS

We will make the WHS data available in several ways:

1. As a complete download (with information about and links for recommended viewing applications)
2. As a web accessible application
3. As downloadable images

We plan to have an initial implementation of a WHS prototype by the Neuroinformatics 2009 Congress. At that time, we will demonstrate our progress to interested members of the digital atlasing community and solicit their feedback, input for future directions, and volunteers for future activities.

2.3.5 Immediate Goals for the Digital Atlasing Task Force

Since we have defined a vision and recommended immediate steps that will aid in moving this field toward this vision, we can focus our efforts on the longer-term infrastructure issues and begin to create more concrete standards and recommendations for which tools and resources to target for specific components of this infrastructure.

Our immediate plans include:

- Develop a set of best-practices for ensuring particular data types will be WHS-compatible and can more easily be brought into this atlasing framework. The first data types will include:
  - MRI
  - Nissl volumes
  - Certain gene expression data, e.g., 2D image slices
- Identify and prioritize implementation components of the longer-term infrastructure goals

2.4 Longer-Term Goals for Interoperable Digital Atlasing Infrastructure

2.4.1 Overview of INCF Digital Atlasing Infrastructure (INCF-DAI)

As discussed earlier, we believe that the section above outlines crucial steps towards facilitating the sharing of international atlas resources. However, to meet the original vision put forward by this group, a larger system needs to be put into place. We have identified the essentials of this infrastructure based on work in other areas and neuroscience researchers' needs. We are in the process of expanding this infrastructure and analyzing current resources to see how they might fit into such an infrastructure, and to consider where key items are missing. This is to identify more specific areas where this program, INCF, and the rest of the community may most effectively play a role in filling these key resources.

a. Requirements

In addition to WHS and the standards that have been identified in the section above, there is additional infrastructure that needs to be built in order to facilitate interoperability between atlases and data sharing. Underlying standards for these atlases must exist; however, these atlases shouldn't be greatly constrained by these standards, instead there should also be a system built that can keep track of the information about the different atlases, and convert information in a tool-accessible manner.

In order to create this infrastructure, an architecture document and a forum for discussion and improvement of this architecture must be created. These will allow members of the community to both contribute their ideas as well as build specific components of this system.

b. Architecture Vision

As discussed earlier, location information across scales and coordinates can be attained directly using atlas coordinates, labels, or by spatial placement rules. At this point, we focus primarily on the first two more direct forms of accessing location information, but still take into consideration spatial placement rules for the future.

The INCF digital atlasing infrastructure (INCF-DAI) is envisioned as a collection of distributed services that support publication, discovery and invocation of heterogeneous atlas resources (Figure 4). It is expected that, in most cases, INCF-

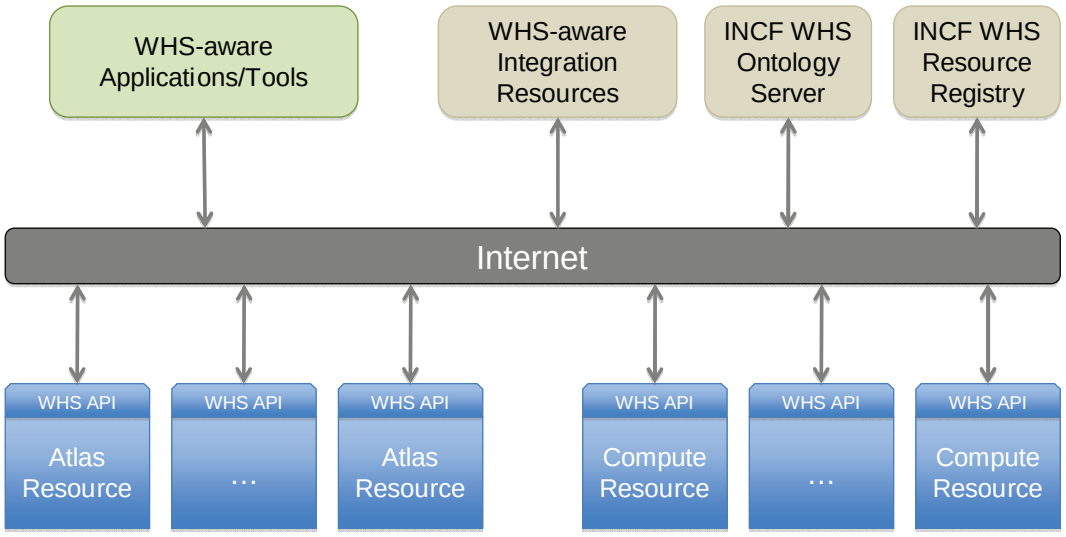


Figure 4: *INCF digital atlasing infrastructure IINCF-DAI) shall be constructed from a collection of distributed services that support publication, discovery, and invocation of heterogeneous atlas resources.*

DAI will reference remote and autonomously supported and updated resources, and host or mirror some resources that are critical for the infrastructure to operate. The resources will include atlasing datasets of different types (shown in Figure 5), compute resources, applications, and workflows. INCF-DAI will establish common access mechanisms for resources of each type, and provide and govern respective standard API development, to ensure syntactic interoperability within the system.

The atlasing integration framework will further rely on common space conventions (WHS) and semantic conventions, each of which will be supported by respective integration servers. By relying on a uniform set of standards and access APIs, INCF-DAI will be scalable in terms of data and compute resources that can be made available within the system.

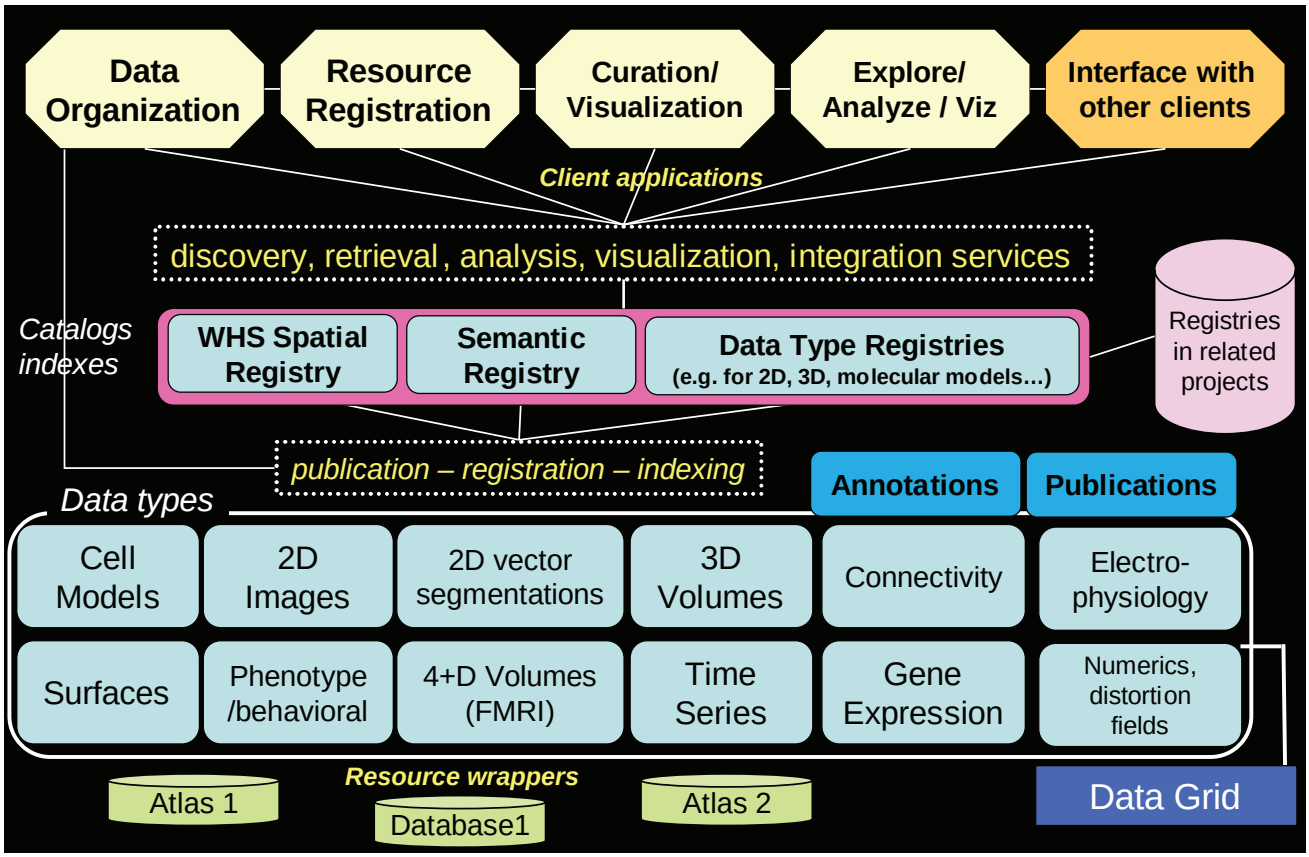
The resources will be registered in an INCF WHS registry. Once registered, each resource will be harvested using the published services, or crawled, to populate an INCF-DAI metadata catalog. We expect that a range of WHS-aware applications will be developed, taking advantage of the INCF-DAI metadata catalog and the common set of services published by each resource. Figure 4 (previous page) captures this vision at a conceptual level.

Figure 5 presents a more detailed depiction of INCF-DAI. It focuses on the data types to be exposed by INCF-DAI resources, and the middleware components that support registration and indexing of different resources. It also demonstrates a sequence of client interfaces that reflect common research workflow components: from data organization, registration and curation to discovery, access, analysis/modeling, publication and communication of results, and WHS-based integration with other atlas interfaces.

INCF-DAI components emphasized in Figure 5 include data types that the infrastructure shall be aware of, the services that manage publication, registration, discovery, query, visualization and integration of atlasing data, and client applications that access INCF-DAI resources. This organization follows a common three-tier approach to service-oriented architecture (SOA) systems. Derived from this organization, the critical backend issues to address in INCF-DAI design are:

1. information modeling and API for each resource type, including canonical relational and XML schemas
2. development of web service wrappers that expose individual sources as one or more resource types with uniform APIs
3. services to convert source- and scale-specific location descriptions to WHS
4. services to convert individual semantic descriptions to INCF-recognized ontologies and supported lexicons
5. population of metadata registries using web service wrappers (*2*) and conversion services (*3 and 4*), to enable

Figure 5: *INCF-DAI resources will include multiple data types that are accessed via common mechanisms for each data type. Services and registries shall be created that (a) manage conversions between the sources and INCF-DAI standards, and (b) expose information to client applications.*



atlas data discovery and integration; synchrony of metadata registries with sources

6. policies and mechanisms for resource curation and quality control

A brief example of a strawman API envisioned for gene expression data that addresses items 1 and 2 is illustrated in Appendix E.

As discussed earlier, spatial and semantic alignments are needed for atlas integration. The INCF-DAI faces interesting challenges with respect to issues 3 and 4 above, due to multiple ways atlas locations can be referenced at different scales, and the diversity and richness of ontology development efforts in the community. Potential solutions include:

- Referencing atlas locations by WHS coordinates:

  This is the common atlas location description adopted in the prior discussion section. However, this location description is limited to scales on which WHS space is fairly easy to quantify, and is problematic at cellular or sub-cellular scales.

  In a simple initial case, a query in WHS coordinates would return properties of an area that is within the object specified by WHS coordinates. In a more general case, one could request a set of properties of an area defined by any spatial relation R (e.g. "contains", "touches", "adjacent", "within N microns") to the object specified by WHS coordinates.

  A technical implementation would include a source wrapper responsible for translating WHS coordinates into the native spatial representation, and processing spatial relationship R as specified in the query.

- Referencing atlas locations by tissue label:

  This location description assumes that tissue label mismatches can be resolved between atlases. A scalable way to do this is to maintain an INCF-approved vocabulary (ontology) of tissue names, and maintain mappings between naming conventions adopted in individual atlases, and the INCF vocabulary. Note that these are not necessarily one to one mappings.

  The location can be specified in the form "WHS:tissue" (where "WHS:" is a prefix indicating that the tissue label comes from WHS-approved vocabulary).

  Again, we envision that a technical implementation would include source wrapper functionality to translate WHS:tissue into some NativeVocabulaty:tissue that the source understands.

- Referencing atlas locations by spatial placement rules:

  WHS-defined ROI specified as a collection of spatial rules that sequentially narrow the ROI space. This type of location description is general enough to be used at scales where structures from the WHS-approved ontology are identified but the WHS coordinate space is poorly defined.

  The location can be specified as a set of rules e.g., "within N microns from WHS:tissue1" and "between WHS:tissue2 and WHS:tissue3" and "adjacent to WHS:tissue4".

  A technical implementation implies that a native source wrapper interprets WHS:tissue as NativeVocabulary:tissue structures, computes areas defined by each rule, and then implements (e.g., in the 2D case) a polygon overlay to derive the ROI from a combination of rules.

In other words, WHS alignment functionality shall be present in both the central WHS server (where translation services handle conversions between the three types of location description), and in the wrappers for individual servers. In the semantic space, the issues are similar, requiring each source to subscribe to a specific vocabulary (ontology), and providing ontology cross-walks and lexical translations at the central INCF ontology server.

c. INCF-DAI Specifications vs. Implementation

The architecture vision outlined above is the first step towards creating a more detailed specifications and architecture document. The development of these will aid in identifying specific areas where this program, INCF, and the rest of the community may most effectively play a role in filling these key resources.

The role of this group is to develop these specifications for INCF-DAI rather than to actually implement the recommended infrastructure. Implementation of this infrastructure is beyond the scope of the INCF digital atlasing program and requires a larger community effort that is backed up by funding by the member country nodes of INCF.

2.4.2 Towards the INCF-DAI Specification

a. Prototype

As discussed earlier, atlas interoperability is to be supported by all three methods above, WHS addresses the first of these methods, but the latter two modes of integration are critical where the variation in brain images are too substantial to allow their registration to the standard WHS coordinate space. They will also play an important role when dealing with the temporal aspects of brain development and the integration of related data sets. In addition to a precise definition of the conceptual Waxholm Space and its realization as a standard canonical atlas, we also propose a project to both test and begin the development of referencing atlas location by tissue label and by spatial placement rules. This project will focus on using WHS to integrate the Edinburgh Mouse Atlas with other developmental atlas data, specifically the Allen Brain Atlas, but time permitting, other atlas data.

The development of conceptual solutions and the specification of systems and infrastructures are greatly assisted by prototypes that can ground the relevant discussions in concrete examples and their implementation. The purpose of this project is two-fold:

1. to develop a prototype for developmental aspects as well as semantics-based integration aspects of Waxholm Space (SemDev prototype)

2. to develop specifications for the architecture and components of the INCF-DAI relating to developmental and semantic integration issues

Again we note the goal here is not to implement the complete INCF-DAI, as this would be substantially beyond the scope of this work, but to ground the development of the specifications of DAI components dealing with developmental data and semantic atlas integration in concrete examples (Figure 6).

We expect that the use of prototype implementations will raise several issues for the specifications and that we will have a number of specification/implementation iterations.

Specifically, the objectives of the project are:

- To demonstrate WHS-based querying of two atlases based on tissue labels (includes issues of mapping between different brain anatomy vocabularies)
- To demonstrate WHS-based querying of two developmental atlases (includes issues of temporal resolution and mapping in WHS)
- To demonstrate WHS-based querying using spatial placement rules (includes spatial reasoning over combination of coordinate space and tissue labels)
- To use all of the above to develop the specification of the INCF -DAI, especially with respect to developmental data and semantic atlas integration. Details will be outlined using the UML language

We note that the specifications of INCF-DAI components emerging from this work on developmental aspects of mouse brain and the semantics-based integration of mouse brain atlases will form an integral part of the wider INCF-DAI specifications, i.e., be based on the same overall system architecture vision.

Also, there is an important synergy required between this task force and the activities of the INCF Program on Ontologies of Neural Structures (PONS). The WHS Atlas will provide specific spatial and temporal modeling requirements that need to be captured by PONS. Similarly, integration of atlases based on anatomical labels and spatial rules will benefit greatly from the deliverables of PONS. The proposed prototype work will thus play a critical role in the collaboration of these two INCF programs.

Although the prototypes are primarily used to drive the specification of the WHS infrastructure, any software products that are deemed to be useful to the general public beyond this primary goal will be made available through the INCF software center. Furthermore, all software developed for the INCF will be available for reuse for anyone on an emerging INCF Digital Atlas Portal.

b. Community Engagement

As in all aspects of this program, it is crucial that the interested parties in the scientific community are involved in multiple steps of this process. This interaction shall be facilitated by INCF and should include both in-person workshops and web tools for on-line feedback and discussion.

2.4.3 Towards INCF-DAI Implementation

Again, the goals of this group are to specify rather than implement INCF-DAI. However, it is both desirable and likely that at least some prototype pieces of this infrastructure will be developed during its specification. Any usable components that are developed will be shared with the community through INCF. As with any complex infrastructure, we expect the implementation of INCF-DAI to be iterative, due to new events and issues that arise during development.

2.4.4 Connecting Key Resources to WHS through INCF-DAI

The connection of key resources to WHS via INCF-DAI serves two purposes. First, it links valuable resources and information to the environment outlined in this document. In addition, the practical goal of linking these through INCF-DAI aids in developing the specifications and standards of the infra-

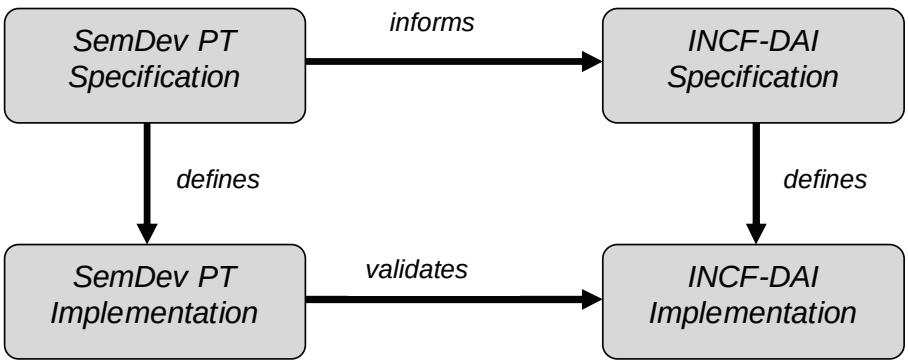


Figure 6: *Relationship between SemDev prototype and INCF-DAI. The primary deliverable is the specification of the INCF-DAI with respect to developmental data and semantic atlas integration, though a simple demonstration of the selected functionality of the SemDev prototype will also be implemented.*

structure. Issues that will likely occur frequently will be highlighted and can be addressed early in development.

The development of neuroinformatics systems has, to date, largely reflected the research and data production goals of individual groups. The variety and scope of these resources are impressive but pose significant challenges for interoperability and standardization. The goal of complete interoperability and standardization is of course infeasible and not necessarily even desirable. However, an outcome of this working group is the identification of key core resources that should:

- Offer basic interfaces to the central WHS standard
- Operate interchangeably with high interconnectivity
- Serve as basic models for connecting new resources

A selection of identified resources is listed below, and a description of the role these may serve in the WHS architecture is discussed in Appendix F. This is not to suggest that these are the only resources that have value in this context but they are neuroinformatics databases that have high usage and utility and they lend themselves to addition to this environment. Further there are resources that will emerge moving forward that should be considered for linkage. While new tools and resources can be built to fit into this infrastructure, the standards and services proposed here can be used by key existing resources to link them to this infrastructure.

A list of some existing key resources that should be considered for early linkage to WHS:

Allen Brain Atlas: www.brain-map.org

EMAGE/EMAP:
genex.hgu.mrc.ac.uk/Emage/database/emageIntro.html

GENSAT: www.gensat.org/index.html

BAMS: brancusi.usc.edu/bkms

NeuroTerrain Atlas: www.neuroterrain.org

Cell Centered Database (CCDB) and SmartAtlas:
ccdb.ucsd.edu/CCDBWebSite/index.html

GeneNetwork: www.genenetwork.org/home.html

Mouse BIRN MRI Atlases (Caltech, Duke, UCLA):
www.nbirn.net/bdr/ bdr_current_data.shtm

Mouse BIRN CIVM Developmental Atlas:
www.civm.duhs.duke.edu/devatlas/index.html

BIRN Mori Mouse DTI Volume Atlas 1.0:
www.nbirn.net/bdr/mouse_atlas_dti/index.shtm

High Resolution Mouse Brain Atlas (Sidman):
www.hms.harvard.edu/research/brain/intro.html

EuroExpress/GenePaint:
www.eurexpress.org/ee_old/technology/publish.html

Paxinos Reference Atlas: www.elsevier.com

An initial analysis of these resources has helped us develop the infrastructure and projects that we have proposed above. The incorporation and adoption of these key resources into the WHS paradigm will require detailed examination of the goals, methods, interfaces, and data sources offered. These core resources will serve as the basis for interchanging data and linking new resources into the WHS standard.

A much more detailed analysis of these and other resources within the context of the INCF-DAI will be required in order to develop the common data models, standards, and other pieces of infrastructure that need to be built, such as APIs and tool packages. A next set of critical path activities is to:

- Perform a detailed inventory of the above key standard resources examining utility, input/output relationships, and research goals. This analysis will be performed with an emphasis on creating an architecture document for INCF-DAI
- Identify linkage maps to the WHS space standard and the set of interchange queries that allow communication between the resources
- Formulate a development plan that enables each resource to be appropriately mapped

We believe that some of this initial work should be completed by the WHS Task Force; however, this will require ongoing specifications of standards, APIs and tool packages which should fall to the WHS and INCF-DAI task forces and other members of the neuroinformatics community.

2.5 Long-Term Infrastructure Implementation Recommendations

To summarize, upon the formation of this task force, we developed a high-level vision for the infrastructure needed to facilitate digital atlasing resources for the international research community. During the time that has followed, we gathered specific use cases to drive our direction and started an analysis of existing resources in order to identify where they fit into this infrastructure as well as to identify where gaps exist in moving towards these goals. Based on this work, we have developed a vision for an extensive and expandable infrastructure for an atlasing framework and identified an obvious need for the generation of a set of standards and APIs.

We have also proposed two projects to quickly make headway on creating some key components of this infrastructure. In addition, these projects will aid us in further defining INCF-DAI. These in conjunction with ongoing work will allow us to create an architecture document for infrastructure components. Once this is created, it will be much easier for community members to focus on contributing to specific components of the INCF-DAI.

While WHS is being built, headway will also be made towards INCF-DAI. We believe this group can continue to lead the development of these resources; however, in order to tackle tasks specific to a community, experts from those groups will be required. In addition, certain tasks will require technical

expertise to help develop, implement, and test the practical incorporation of standards and APIs into usable software.

We recommend that this group continue and focus on the following issues, while ad-hoc committees are formed to focus on developing specific components of this environment.

2.5.1 Long-Term Efforts by this Task Force

Once we have identified and prioritized the implementation components of the longer-term infrastructure goals, the next steps are to:

- Identify existing resources (tools and services) that may be used in the INCF-DAI
- Examine current common data models and exchange standards for targeted data types:
  - If none exists, encourage the appropriate community to create one
  - Use these to create APIs and source wrappers
  - Develop a global data model that crosses data types
- Identify the appropriate people and work with them when appropriate to create:
  - Standards
  - APIs
  - Software tools
  - Services and registries
- Facilitate community engagement:
  - Disseminate information via publications, workshops, and seminars
  - Gather feedback to improve the Atlasing program

2.5.2 Recommended Ad-hoc Committees

WHS Task Force

- As described earlier in the document, this group will focus primarily on refining and building WHS

Registration Task Force

- In order for the system to be accessible to researchers, several issues related to spatial registration will need special focus. This group should thoroughly analyze, test, and create recommendations and standards in this area

Common Data Model Task Force(s)

- For each targeted data type, external community specialists will be needed to develop support for their data type

Digital INCF-DAI Task Force

- This group will create an architecture document for the INCF infrastructure components:
  - They will ensure that the standards, services, registries, and APIs are usable for tool builders
  - They will facilitate distribution of these resources to other tool-builders as they are developed

In addition, each of these committees will participate in dissemination of their findings and products to the community and gather feedback for improvement.

2.6 Why the INCF?

Standards, guidelines, infrastructure, and related tools are rarely successful by mandate but rather through compelling utility and broad exposure. The efforts of the W3C (www.w3.org) are testimony to this model and the INCF has initiated this objective for development of neuroinformatics methods in the life sciences. The role of the INCF in the proposed vision may be seen to be fully consistent with the continuing objectives of the organization both as a promoter and supporter of the work. The proposed system will require significant architecture and software engineering, particularly in the initial stages but the major pieces and technologies are already available.

The INCF is positioned in a unique position to facilitate this sort of international collaborative activity. It is clear that the infrastructure pieces must be encouraged to be built by the informatics members of the scientific community and by the individual resource stakeholders with guidance and support from INCF. Ultimately as the system develops and later matures there will be more opportunity for the integration of new resources. The INCF may act as a clearinghouse for these technologies and datasets to enable the system to maintain a living character. Specifically, the role of the INCF may initially be to:

- Promote the proposed vision as it is scientifically vetted
- Encourage the architectural team that will specify the detailed design and connection of key resources to the WHS standard space by organizing meetings and discussion groups to maintain the momentum to complete the task at hand
- Monitor the development of the software development and provide connection between relevant parties in that effort
- Assist in finding a host site for the system and help to maintain and promote its use within the community
- Facilitate the system if successful as a standard operating procedure in rodent atlasing

3 References

Baldock RA, et al. (2003) EMAP and EMAGE: A framework for understanding spatially organized data. Neuroinformatics, 1: 309-25.

Bello M, et al. (2005) Hybrid segmentation framework for tissue images containing gene expression data. Presented at MICCAI.

Bertrand L, Nissanov J (2008) The Neuroterrain 3D Mouse Brain Atlas. Front Neuroinformatics, 2:3. doi:10.3389/neuro.11.003.2008.

Bjaalie JG (2002) Localization in the brain: new solutions emerging. Nature Reviews Neuroscience, 3: 322-5.

Blake JA, et al. (2003) MGD: the Mouse Genome Database, Nucleic Acids Research, 31:193-5.

Boline J, et al. (2007) Workshop report: 1st INCF Workshop on Mouse and Rat Brain Digital Atlasing Systems. Available from Nature Precedings: dx.doi.org/10.1038/npre.2007.1046.1

Boline J, et al. (2008) Digital atlases as a framework for data sharing. Front Neurosci, 2:100-106. doi:10.3389/neuro.01.012.2008.

Carson JP, et al. (2005) A method for automated detection of gene expression required for the establishment of a digital transcriptome-wide gene expression atlas. J Microscopy, 217:275-81.

Carson JP, et al. (2005) A Digital Atlas to Characterize the Mouse Brain Transcriptome. PLoS Comp Biol, 1:e41.

Fitzpatrick JM, et al. (2000) Image Registration, in Handbook of Medical Imaging: Medical Image Processing and Analysis, Vol. 2. M Sonka & JM Fitzpatrick, Eds. SPIE Press, 447-514.

Hajnal JV, et al. (2001) Medical Image Registration: CRC Press.

Herzig U, et al. (2001) Development of high-throughput tools to unravel the complexity of gene expression patterns in the mammalian brain. Presented at Symposium on Complexity in Biological Information Processing, Berlin, Germany.

Jahne B (2002) Digital Image Processing, 5th ed. Berlin: Springer.

Johnson GA, et al. (2007) High-throughput morphologic phenotyping of the mouse brain with magnetic resonance histology. NeuroImage, 37:82-9.

Ju T, et al. (2003) A geometric database for gene expression data. Presented at Symposium on Geometry Processing, Aachen, Germany.

Kakadiaris IA, et al. (2004) Landmark-driven, atlas-based segmentation of mouse brain tissue images containing gene expression data. Medical Image Computing and Computer-Assisted Intervention: 7th International Conference, St-Malo, France. Proceedings, Part I. 3216:192-9.

Koslow S, Subramanian S (2005) Databasing the Brain: From Data to Knowledge (Neuroinformatics), John Wiley & Sons.

MacKenzie-Graham A, et al. (2003) The informatics of a C57BL/6J mouse brain atlas. Neuroinformatics, 1:397-410.

Mattes D, et al. (2003) PET-CT image registration in the chest using free-form deformations. IEEE Trans Medical Imaging, 22:120.

Martone ME, et al. (2004) E-neuroscience: challenges and triumphs in integrating distributed data from molecules to brains. Nat Neurosci, 7:467-72.

Ng L, et al. (2003) Neuroinformatics for genomewide 3-D gene expression mapping in the mouse brain. IEEE Trans Comp Biol Bioinformatics, 4:382-93.

Pluim JPW, et al. (2003) Mutual information-based registration of medical images: a survey. IEEE Trans Medical Imaging, 22:986.

Raya SP, Udupa JK (1990) Shape-Based Interpolation of Multidimensional Objects. IEEE Trans Medical Imaging, 9:32-42.

Rex DE, et al. (2003) The LONI Pipeline Processing Environment. Neuroimage, 19:1033-48.

Schwartz EL, et al. (1998) Applications of computer-graphics and image processing to 2D and 3D modeling of the functional architecture of visual cortex. IEEE Comp Graphics Applications, 8:13-23.

Sunkin S (2006) Towards integration of murine spatial resolution expression databases. Trends Genetics, 22:211-7.

Toga AW (2002) Neuroimage databases: The good, the bad and the ugly. Nat Rev Neurosci, 4:302-9.

Toga AW (2002) Imaging databases and neuroscience. Neuroscientist, 8:423-36.

Toga AW, Mazziotta JC (1996) Brain Mapping: The Methods. San Diego: Academic Press.

Viola P, Wells WM (1997) Alignment by maximization of mutual information. Int J Comp Vision, 24:137-54.

Visel A, et al. (2002) A gene expression map of the mouse brain: Genepaint.org - A database of gene expression patterns, in Neuroscience Databases: A Practical Guide, R. Kotter, Ed. Boston, 19-36.

Visel A, et al. (2004) GenePaint.org: An atlas of gene expression patterns in the mouse embryo. Nucleic Acids Res, 32:D552-6.

Warren J, Weimer H (2002) Subdivision meshes for geometric design: a constructive approach. San Francisco: Morgan Kaufman Publishers.

Yushkevich PA, et al. (2005) Using MRI to build a 3D reference atlas of the mouse brain from histology images. Presented at Intl Soc Magn Res Med.

Zreiqat H, et al. (1998) Quantitative aspects of an in situ hybridization procedure for detecting mRNAs in cells using 96-well microplates. Mol Biotechnology, 10:107-13.

Appendix A: Terms

User:

We refer to a "user" as a biological scientist, tool-developer, or group that might want to access or use any component of the system.

A user may wish to contribute data at different levels, including individual datasets, atlases, large sets of data, complete databases, or atlases with related meta-data.

Reference atlas:

We refer to a reference atlas as a digital atlas that has been brought into Waxholm Space. The atlas along with its registration transformations required to move it into Waxholm Space is available to external groups.

The following terms and definitions are taken from Boline J, et al (2007) Workshop report: 1st INCF Workshop on Mouse and Rat Brain Digital Atlasing Systems. Available from Nature Precedings: dx.doi.org/10.1038/npre.2007.1046.1

Digital Atlas:

An atlas is a collection of maps or manifolds, traditionally bound into book form, but also found in multimedia formats (*Wikipedia (3/20/07)*. As technology has advanced so have brain atlases transformed from passive paper guides to dynamic databases at the core of software applications (Toga 2002). In this report, we almost exclusively focus on these sophisticated digital atlases held in either free-standing software tools or web-enabled hyperlinked neuroinformatics hubs that act a gateway to a collection of databases, metadata catalogs, and related multimedia documents and annotations that are placed in a common spatial framework and thus can be juxtaposed and analyzed together.

Data Repository:

A central place where data is stored and maintained. (*Wikipedia 04/07)*

Database:

A database can be defined as a structured collection of records or data that is stored in a computer so that a program can consult it to answer queries and incorporates software to make it accessible in a variety of ways. The records retrieved in answer to queries become information that can be used to make decisions. (*Adapted from Wikipedia 04/26/07 and the Oxford English Dictionary)*

Spatial Database:

"We propose a definition of a spatial database system as a database system that offers spatial data types in its data model and query language and supports spatial data types in its implementation, providing at least spatial indexing and spatial join methods." (*Ralf Güting, www.informatik.fernunihagen.de/import/pi4/papers/IntroSpatialDBMS.pdf*

Image Registration:

Image registration is a process of relating and organizing characteristics of two or more images, so that image data obtained from different measurements can be discovered, compared or integrated. Image registration may include organizing image metadata into an image metadata catalog, as well as placing the images in a common semantic or spatial framework. This may be contrasted with semantic registration that involves verifying image metadata and associated labeled delineations against established formal ontologies or controlled vocabularies, to ensure commonality of terms used in image description.

Spatial registration (often also called "image registration") is the process of modifying spatial characteristics of an image dataset to align it to another image dataset, thus placing different images into a common coordinate reference frame. Image registration techniques vary across domains. For example, different MR images may be spatially co-registered by tuning image metadata which includes pixel size, dimensions and orientation of the image. Alternately, an image may be transformed into alignment with another image by specifying pairs of fiducial control points, or by relating the image to a set of anatomic feature delineations. Technically, spatial registration procedures may involve a linear alignment or a nonlinear alignment, which actually warps the images into a common space. Spatial registration ensures that images may be discovered and queried by spatial coordinates, via an anatomic atlas.

Spatial Registry or Spatial Registration Services:

Spatial registry is a component of image metadata catalog. It contains information about position, orientation and extent of registered images, and links image spatial metadata with other image metadata. A spatial registry is typically organized as a spatial database: it contains polygonal representations of registered images, maintains spatial indexing of the image polygons, and supports spatial queries, e.g. 'select images intersecting with a user-defined shape,' 'select images whose centroids are contained within a user-defined shape,' 'select images found within a 3mm sphere around a user-defined point.' The key concept is that the spatial registry connects image data with data annotation enabling effect inquiry.

Annotation:

Annotation is the process of associating information content and knowledge with raw data. Annotations may differ in purpose and complexity, ranging from simple text notes made at a particular point in a document or in an atlas, to multimedia composite objects that may include user-defined shapes, documents, hyperlinks, or other annotations.

Metadata:

Metadata is data about or associated with data used to render a more precise description or record of its significance. An item of metadata may describe an individual data item or a collection of data items and is used to facilitate the understanding, use and management of data. (*Adapted from Wikipedia)*

Application Programming Interface (API):

An API is a source code interface that a computer system or program library provides in order to support requests for services to be made of it by a computer program. *(Wikipedia 04/07)*

Appendix B: Best Practices

Standardizing Interoperability of MR/ Histology Atlases of the Mouse Brain

Introduction

Over the last ten years there has been increasing effort to produce digital atlases of MR and conventional histology images of the mouse brain. Each atlasing group has developed their own approach and standards. There is a compelling need to rationalize these efforts to:

- make the data from all the groups more uniformly available to the neuroimaging community
- allow comparison between atlases
- provide a framework that would encourage cooperation in future efforts.

The "best practices" that we will describe are not necessarily the "best" practices and certainly are not the only approach. Rather, this will provide a starting point-a point from which we fully expect more sophisticated and complete approaches will be developed.

Background

Magnetic resonance microscopy (MRM) provides a unique tool to facilitate comparisons of brains and brain atlases. MRM is nondestructive so the intact tissue is completely undistorted by shrinkage or cutting. Data can be visualized along any axis without distortion or loss of spatial resolution (Figure B.1 a and b). But as with conventional optical histology, there are an abundance of imaging methods each producing unique tissue contrast. Figure B.1 b and c shows representative images from the same specimen with two differing sets of acquisition parameters. Layers in the hippocampus seen in b are not apparent in c. But cortical layering and subthalamic nuclei in c are not evident in b. The widely varying contrast in the cytoarchitecture is clear. In order for registration software to work effectively, there must be some common structures visible in the target and sample data sets.

Goal

The global goal of the best practices document is to provide methodology that will facilitate comparison of digital atlases of MR and conventional histology. More specifically we propose the following specific aims:

- We will articulate a common set of methods that will allow groups to prepare digital MR data with minimal distortion and common contrast so that existing algorithms will permit registration based on maximization of mutual information.
- We will define an approach that will allow registration of digital MR data to conventional Nissl histology.
- We will provide a canonical atlas of MR and conventional histology using the protocols we have defined in this document.
- We will register this canonical MR/Nissl set to at least two labeled atlases (Allen Brain Atlas, LONI/MBAT) to provide a common ontology and links to tools for gene expression.
- We will establish and validate the protocols required to use this standardized approach as a method to map, MR, conventional optical, and traditional labeled atlases into Waxholm space.

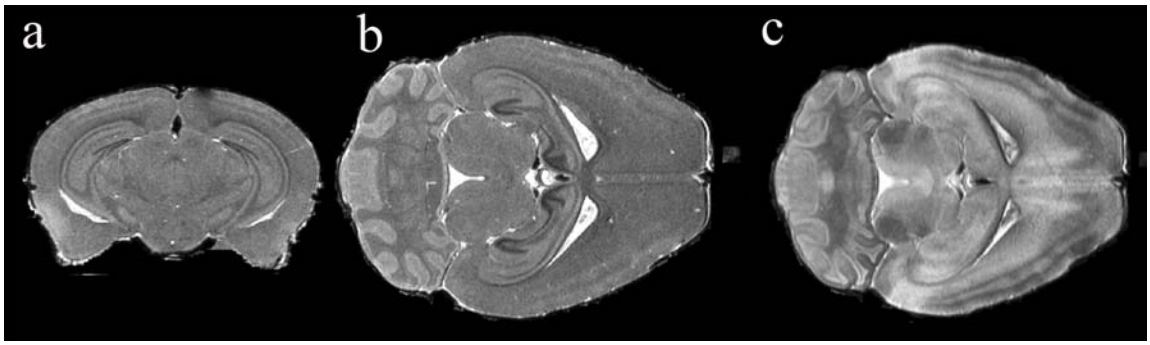


Figure B.1: *a) Coronal section from a 3D MR microscopy image of an actively stained mouse brain; b) dorsal plane from the same data sets showing isotropic resolution. Since the brain is fixed in the skull there is limited shrinkage and no physical distortion from sectioning. c) Dorsal plane of the very same specimen acquired with different acquisition protocol demonstrates completely different cytoarchitecture.*

Appendix C: Use Cases

C.1 Biologists: Analysis Environment (few time points, within species)

Toxicologist-Pharmacologist

Atlas as Reference (traditional atlas use expanded into genomic analysis)

User Profile: I am a toxicologist/pharmacologist with basic knowledge of mammalian brain anatomy. I am looking at the long-term impact of a drug to potentially treat seizures in young children.

Usage Scenario: I have series of sections from a set of controls and an animal treated with the pharmacological agent as a neonate. The experimental and control animals are sacrificed for gross anatomical analysis and cell morphology followed by gene expression analysis. The phenotype in the experimental animals is broader than expected and encompasses brain regions in which I am not well-versed. The control littermates' brains are clearly normal but I need to identify the affected regions in my experimental animals. I load images from the experimental brains into the atlas workspace to compare with the labeled atlas. In the comparison view, I identify the affected regions on the atlas, and am able to compute area and volume changes in these affected regions. I then query the atlas for genes expressed in the regions of interest. With this information, I select region-specific probes to perform in situ hybridizations, quantitative PCR and microarray studies of the treated animals.

What I will gain: An atlas tool which links anatomy, gene expression data, literature and other spatially and semantically mapped data will provide clues for the next experiments. The visual nature of atlases provides a mechanism to avoid arguments of nomenclature and convey observations and ideas.

Neuroscientist/Geneticist

User Profile: I am a neuroscientist with a mouse model of autism. We have extensive behavioral and genetic information. We believe there are major differences in brain morphometry.

Use Scenario: I would like to compare the morphometry of the brain in our model to a standard model (C57BL/) of SV 129. More specifically I would like to get the volume of as many specific regions of the brain as possible to identify areas where we might focus our interest.

I would like a standardized scanning protocol that would produce a dataset that can be related to the work of others. I do not have MRI experience so I would like some help interpreting the data. We don't have MRI equipment so I need access to a scanning system. I am willing to submit our images to a database for others to use after we have published our findings.

What I will gain: I am very interested in MRI but am confused by the wide variety of information I see in the literature. I can see some of the structures of interest in some publications, but not in others. With a standardized protocol, I am hopeful that I can compare changes in our model with those of others. I would like some sort of uniform acquisition parameters and a set of display tools to make this possible.

Toxicologist

User Profile: I am a toxicologist at a pharmaceutical company. We have a compound that shows great promise for treatment of prostate cancer. But behavioral tests suggest there may be some neurotoxicity. We would like to know if the brains in our Fisher 344 rats have undergone any morphologic change after a full treatment regimen. More specifically, we would like to know the volumes of several brain structures (amygdala, hippocampus, thalamus) and how the volumes change with age.

Use Scenario: I would like to compare my data to that acquired by others. I will need help with data acquisition and reference to a standard atlas so that I know how to identify the regions of interest. I would really like to know about age-related structural changes in these regions so that I can quantify any changes that may be associated with toxic effects from the treatment.

What I will gain: Histology measures can be imprecise. We cannot get volumes, so measures of neuronal loss can only be seen using very time-consuming stereology. However, we don't know the range of normal variation. We would like to pool our data with others. We're willing to work with other pharmaceutical companies as long as we don't have to disclose the specifics of our compounds.

Microarray Experimentalist

User Profile: I am a research scientist investigating a disease model of Parkinson's disease with microarray techniques. I have a strong background in data analysis, but little experience in computation and informatics. I use my desktop computer at work, which is connected to our university's internet.

Usage Scenario: I want to share a set of microarray data that has been generated using punches of different brain areas from groups of control and of diseased mice, as well as a group of animals that have been treated with a therapy that seems to slow the onset of Parkinsonian symptoms. As I am ready to publish these findings in a journal, I would like to share the data in such a manner that other researchers can visually examine the results and compare the group differences. Also, I would like to examine the genes that vary most among the different groups, and visualize the associated brain regions. Of course, impressive images for the journal article would be great!

I am willing to submit either my raw or processed data, both are currently held in separate very large Excel spreadsheets,

I also have the associated annotation information for the microarray that I used and am willing to share that if needed. I have processed the dataset using a commonly used normalization method.

I would like to simply upload my raw and/or processed data and answer some simple questions about my experimental design and the subjects in the experiment. I would like to register the spatial location of the punches to an atlas, and then I would like to use the atlas to visualize and compare the expression of different genes in relation to the punches in the different groups of mice.

Others interested in this data might also be examining Parkinson's disease or simply interested in how genes are expressed in different portions of the brain.

What I will gain: With simple tools that allow me to upload and share my data with fellow researchers, we can more rigorously examine the gene expression differences in these groups. We may gain additional insight into some of the causes of Parkinson's, and what might be involved in slowing its progression.

C.2 Biologists: Analysis Environment (across time and/or species)

Developmental Neuroanatomist

User Profile: I am an experimental neuroanatomist devoted to the analysis of connectional organization of neural networks. My laboratory's desktop computers are connected to the internet via a local network.

Usage Scenario: Current research interest of the neuroanatomist is on mechanisms underlying development and maturation of neural circuits. Developmentally drastic and dynamic changes occur in the cerebellar neural circuits during the period of first through third postnatal week. The molecular mechanisms underlying such events are largely unknown. Since, several molecules and genes are considered to be involved in this event, the digital atlas must give full play to acquire comprehensive knowledge useful for exploration of which genes are responsible for which process of the events and how. Morphological examination can be followed and carried out in particular cells and in particular timing of the developments. In this connection, the gene expression database in the digital atlas for a variety of cerebellar neurons will help us to pick candidate genes. Such combination of morphological and molecular biological analysis can only be established effectively by use of the digital atlas, and such application will be expanded in other brain regions and different research questions as well.

What I will gain: By use of the atlas, my research subjects will be much more focused, saving time for data analysis. By adding the analyzed data to the atlas, the atlas itself will be further expanded for use in the scientific community.

Developmental Biologist

User Profile: I am a development biologist trying to understanding when cells within the developing brain are restricted to region specific fates. I also want to know the course of molecular events which bring cells to their differentiated state.

Usage Scenario: A set of developmental brain atlases have been constructed staged by the emergence of anatomical features and the expression of a handful of well-characterized developmentally regulated genes. Presumably, detailed fate maps of the developing brain reliably predict each adult structure's earliest embryonic origin. The fate map resides in a developmental ontology. Stage specific gene expression patterns and the distribution of macromolecules have been, and are continually being spatially and semantically mapped to the developmental atlases. An AGEA like correlation maps can be generated allowing the user to examine the constellation of genes and macromolecules at any developmental stage.

Selecting a region within any of the atlases links the user to predecessors and decedents of that region through the developmental ontology. The stage specific mapping of gene expression and macromolecule distribution can be used to examine the progressive specification of brain regions.

What I will gain: Adding the temporal component to an analysis provides a dynamic view of biological systems. Due to complexity, causality can be difficult to determine; however, correlated sequences of events point toward potentially causal events.

Collaborative Project (across stage and species)

User Profile: I am a researcher, with knowledge in imaging and computing, and I am an experienced user of atlas resources. My research concerns the development and molecular mechanisms of Alzheimer's disease. My project is a collaboration with several molecular biologists, neurobiologists and neuroanatomists. I am not a computer programmer.

Scenario: I have recently observed the progress of the disease in an adult mouse model. From molecular studies I have gained understanding that some genes that are involved in the development of the disease. I want to study these genes in a zebrafish model; an Alzheimer zebrafish is available.

1. Share something: My current findings in mouse; these are raw data (images) and the processed results.
2. Find something: With the morphologists we have clearly identified the regions of the brain in which the genes are expressed. These regions have clear counterparts in mouse brain morphology and are named unambiguously. What are the related parts in the zebrafish brain? Can I study the progression of these parts by looking at development? Is the relation in the developmental sense known?

3. Analyze something: Can I learn how my gene is related to development? From the histological atlas, are there links available to other modalities and data that can help me find this out?

What I will gain: Understanding of the relation of the ontologies with other resources and a clear insight in the relations that exist in the ontologies so that I can understand the hierarchical structure. Also, I need to make sure that the names, parts, and concepts between the model systems are shared, though the zebrafish is to some extent very different from the mouse.

Developmental Neuroanatomist

User Profile: I am an experimental neuroanatomist devoted to the analysis of morphological and developmental regionalization of rodent and other brains, including human. I am a producer of neuroanatomic ontologies, particularly those with a developmental basis. My laboratory's desktop computers are connected to the internet via local network.

Usage Scenario: My current research interests are on mechanisms underlying progressive patterning, differential specification and histogenetic regionalization of vertebrate brains (shared patterns as well as non-shared aspects are equally of interest; normally variables are interpreted qualitatively and not quantitatively, though the latter is possible). This line of research compares fate mapping data at diverse developmental time points up to the adult structure with gene and protein expression patterns indicative of patterning and differentiation (down to cellular level), as well as with analysis of growth and guidance of developing connections in terms of the molecular background. Such a combination of morphological and molecular biological analysis is served effectively by use of digital atlases, but results of this research line frequently lead either to discovery of novelty not contemplated in current atlases (new subdivisions, novel developmental descriptions, new connective data) or to discovery of errors in current atlases. Essentially this means being able to propose changes to current ontologies and to the tracing of boundaries and naming of structures. I would like to see a reasonable way to update atlases to reflect new results as well as to correct earlier errors, with appropriate editorial controls.

What I will gain: By use of continuously corrected atlases and ontologies, the research of all neuroanatomists and developmental biologists will be much better focused, saving time and reducing confusion that might result from due incorect or incomplete interpretations in data analysis, as well as in the design of new experiments.

C.3 Share a Standard Atlas

Neuroimaging Biologist (multiple MR)

User Profile: I am a senior neurobiologist whose lab wants to study the relationship between structural connectivity in the brain with gene expression or other modalities. I have access to reasonably sophisticated computational staff who can successfully navigate the INCF canonical atlas interface.

Usage Scenario: Our lab's 7T magnet and DTI sequence generate high resolution 100 μm^3 voxels to produce brain-wide maps of white matter tracts, providing an unprecedented view of structural connectivity in the brain. We also have high resolution MR images that can be registered to the canonical atlases in the INCF Portal. This enables us to cross-map with other datasets, including gene expression atlases, such as the Allen Brain Atlas and Genepaint, both of which (in this scenario) would have been previously registered to the INCF atlasing infrastructure. After performing the registration, structural connectivity-based queries performed on the DTI atlas can be translated to equivalent queries in other modalities.

What I will gain: The laboratory has access to very high resolution data in one modality and can now cross-map to other resources. This enhances the value of the specialized data and provides connectivity with multiple biological databases and resources.

Neuroanatomist

User Profile: I am a neuroanatomist working with neuroinformatics staff. I have expertise in rodent neuroanatomy with some novel ideas about structural and functional partitions of the mouse or rat brain. I'm not advanced in computational or mapping techniques, but have access to this skill set in my group.

Usage Scenario: I wish to cross-compare annotations drawn in 2D based on histology with other atlases, and to perform a computational study of structural volumes, borders, etc. with target canonical atlases. My team can upload 2D histology data in a prescribed format to the INCF portal which can propose a 3D reconstruction of the annotations. The team will review and accept this, or offer its own reconstruction. We can also supply an MRI of the brain to help guide the reconstruction. From here, annotations can be projected onto stored canonical atlases for cross-comparison, statistical measurement, and study. I should be able to do most of this work myselfin a semi-automated way.

What I will gain: I have the ability to make the new annotations public prior to publication, and to allow others to view these data in the context of canonical atlases. Annotations and comments can be attached to the structural delineations to allow interpretation and discussion.

Histology Atlas Provider

User Profile: I am from a research organization that has created an annotated, histology-based atlas and would like to cross-link our atlas with other atlasing frameworks. Our group has expertise in image registration, informatics and delivering high-resolution images via web-based applications. The online resources associated with the Allen Reference Atlas are potential examples of such a use case: www.brain-map.org/mouse/atlas/ARA/Coronal/browser.html and www.brain-map.org/mouse/atlas/coronal/legend.html

Usage Scenario: Our organization has existing web services which allow a user to browse and navigate the atlas and associated histology at high resolution. We would like to enable interoperability between the atlasing framework and our atlas.

As part of the atlas construction, a 3D reconstruction of the histology images was performed resulting in a consistent 3D histological volume. The first step is to register our histological volume with the canonical atlas. This will involve deformable and possible cross-modality 3D registration. This could be achieved via either in-house resources or with registration tools provided by the atlasing framework.

Once our 3D atlas volume has been registered, we develop interface code to hook up to the framework API. The main goal is to allow users to atlasing framework to seamlessly query and retrieve images, structural ontology and other metadata directly from our site.

What I will gain: Linking up our atlas with the framework will potentially foster dissemination and attract a wider audience. The integration of our atlas also allows us to readily use the resources of other research groups and gives us the potential to make cross-server discoveries.

C.4 Other Types of Resource Providers

Gene Expression Image Provider

User Profile: I am from a research organization which has created a large-scale database of gene expression images of the mouse brain and would like to cross-link our data with the atlasing framework. Our group has expertise in image registration, informatics and delivering high resolution image via web based applications. The Allen Brain Atlas, GenePaint and GENSAT are concrete examples of such a use case.

Usage Scenario: Using the Allen Brain Atlas (ABA) as an example: a set of APIs exist to allow an external programmer to retrieve information for each gene and its associated image series. Each individual image (~1 micron pixel resolution) can be accessed via a web service. The service allows the programmer to specify the region of interest and zoom level and retrieve the resulting jpeg image from the server. Additionally, every ISH image is also registered to the standard 3D reference atlas volume. Thus, by co-registering the ABA atlas to the canonical atlas of the framework it is possible to enable interoperation between the atlasing framework and the ABA image database.

What I will gain: The development of the interface between the atlas framework and the ISH image database foster dissemination and attract a wider user base.

Registration Algorithm Provider

User Profile: I am a neuroscientist who knows how to program and have developed a workflow and set of tools that I would like to share with others along with some of my data.

- I collect coronal Nissl high-resolution digital image slices through the midbrain.
- My workflow and series of scripts allow me to semi-automatically process these images and align them to each other to create a 3D volume of the sub-section of the brain.
- My tools also include the ability to analyze similar volumes from different animals and look for intensity differences in the slices.
- I have some experience with informatics, but primarily deal with my data locally on my lab's fairly small computational network; my algorithms can take some time to run. I have a moderately fast internet connection through the university.

Usage Scenario: While I would like to share my raw and processed data in relation to an atlas, it's more important to me to share my workflow and analysis tools and scripts. I would like to see these tools used, tested, and improved by others. In addition, I would like to see the pool of data analyzed in this manner increase over time so it's easier for me to compare my data to that of other groups.

I would like to share these workflows in a simple manner, by simply uploading it to a website along with some minimal information and documentation. However, since these tools create a 3D volume of the sub-section of the brain, I would like to integrate these tools in such a manner that they can be used to allow a user to register these to a canonical atlas. I am willing to make changes necessary in order to make these tools compatible with these canonical atlases under the following circumstances:

1. I can find out how to do this easily, e.g. with a module that can either be implemented into my tools or for which I can create a simple converter.
2. I would like to integrate something flexible enough where a user can choose whichever reference atlas they wish to use.
3. I can easily alert potential users about this new functionality in my tools.
4. Ideally there is a space provided that would allow feedback from my users, both about their interaction with my tools and the results they obtain with my tools.

What I will gain: Going through this process will give my tools additional visibility. As more datasets are put through the workflow, a larger pool of experiments will be available to me and others for cross-experimental analyses.

Registration Algorithm Provider

User Profile: I am an expert in image processing with a concentration in registration algorithms. I am part of a lab with access to multiple data sets and wish to test the performance of a new proposed algorithm for registration of histological data. I have advanced understanding of software programming and methodology.

Usage Scenario: I spend most of my time working with codes and algorithms, and have a new method for registering a nuclear stain series, possibly Nissl, across modality to a myelin series. I wish to test the performance of this algorithm against canonical datasets and, if successful, propose the algorithm for inclusion in the INCF canonical resource methods catalog. The INCF portal would require in this case the I/O of the algorithm and its API conform to standards that could be published consistently with existing algorithms. I am able to upload an algorithm to run back- and regression-testing against the existing canonical algorithms in the portal. The portal is set up to allow this type of testing and can report performance statistics on the performance of my algorithm against the standards. If the statistics are favorable, I may enter the algorithm into the public resource.

What I will gain: I gain the ability to understand the performance of my algorithm compared to other canonical algorithms and datasets in the INCF portal. This may help drive decisions in algorithm enhancement and lead to publication. The sequence may also uncover weaknesses in the portal's mapping infrastructure that a new algorithm could improve.

Analysis Provider

User Profile: I am from a research organization which has created a set of spatial correlation maps based on gene expression images and wish to share them with the neuroscientific community. The AGEA (Anatomic Gene Expression Atlas) project at the Allen Institute for Brain Science is a concrete example of such a use case: www.brain-map.org/agea/all_coronal.

Usage Scenario: AGEA characterizes the multi-scale spatial relationship in the mouse brain as derived from gene expression data without a prior knowledge of classical anatomy. AGEA is based on over 4000 image series of the Allen Brain Atlas. Each ISH image series is processed through an automated pipeline that detects expressing cells. These cells are then mapped to a standard coordinate system. The 3D reference volume into subdivided into 200 micron voxels. At each voxel, an expression energy number is computed representing the product of expression density and intensity. Pearson's correlation coefficients are computed for each voxel's relation to every other voxel, resulting in over 53 000 3D correlation maps. The correlation values can be interpreted as measures of co-expression between two voxels providing information from mesoscale co-expression among brain regions to gene-based organization of local scale structures. Such maps are of potential scientific interest for understanding structural-functional relationships.

A web-based interface has been developed to allow easy navigation the more than 53 000 maps. The ability to go to a specific map and location using URL parameters allows easy integration with the atlasing framework.

What I will gain: By incorporating a large number of genes, the AGEA correlation maps provide insight into a transcriptome organization of the mouse brain. The cross-integration of AGEA with information from other modalities in the atlas framework such as functional imaging, angiography, diffusion tensor imaging to name a few allows potential for content discovery for a better understanding of brain function and organization.

Atlas Interoperability (spatial mapping across species/genes)

User Profile: I am a developmental biologist studying spatially mapped gene expression patterns during mouse brain development.

Usage Scenario: Until recently, gene expression queries on the EMAGE (Edinburgh Mouse Atlas Gene Expression) database were restricted to one developmental stage per query. For a subset of EMAGE it is now possible to spatially search multiple developmental stages in one query. This has been made possible by adding between 60 and 90 points on the wholemount models which are used to establish roughly equivalent spatial regions across time. Similar transformations within other, related atlases, and spatial mappings across atlases enable the integration of spatial gene expression analysis across multiple developmental stages. In, principle, one could envision this mechanism to cross species boundaries, supporting comparative analysis of brain development.

What I will gain: Temporal aspects of development (and possibly disease) can be studied using spatially-mapped data, complementing an ontology-based approach.

C.5 Data User

Computer Modeler

User Profile: I am a computer scientist modeler who works on creating algorithms to look for trends across multiple diverse datasets in relation to several experimental variables. I have access to a supercomputer and an Internet2 connection.

Usage Scenario: I have developed some new algorithms to examine multiple 3D bitmap datasets (voxel data) for correlations (both negative and positive) and trends across datasets in relation to several experimental variables. It can be used across modalities, subjects, etc., as long as the data are registered to the same "space." Results identify different "locations"

of interest that show variability in a population and potential variables that may be involved in the variability in that location.

I wish to continue testing these algorithms with a large and diverse set of data from different locations and modalities. It is possible that I will find biologically interesting results from these tests, but at this point I just wish to begin analyzing the datasets.

I would like to easily find all 3D data that have been registered to a canonical atlas regardless of the type of experiment. I will need access to the processed datasets (since my algorithms need data that are in the same "space") and as much information as possible about the experiments. preferably in a manner that makes it easy to compare among the different experiments, e.g. a table containing information from a controlled vocabulary. I assume these datasets will need to be moved to a location that is accessible to my supercomputing institute, so it's likely I'll need to download everything.

When I'm ready to share my results, I'd like to link the information to one of the canonical atlases and allow a user to easily visualize the information.

What I will gain: This gives me access to a large and diverse dataset, which is difficult to find. It allows me to continue testing my algorithms on all the data as well as its subsets (I may even be able to test the results based on some of the contributed experiments), and I hope to improve my algorithms based on these findings.

Appendix D: Tools Analysis Survey to Date

D.1 Data Resources

Summary

The available data resources with significant rodent material consist of 3 types: genomic, literature, and anatomical. The first is by far the largest and includes general genomics and expression databases. The latter is confined, due to limits of microarray technology, to gross anatomical grain. The literature resources include connectivity, transmitter, receptor, behavioral, as well as other information. The anatomical can be subdivided into reference atlases and 'experimental' datasets, though the distinction is at times fuzzy. There are numerous atlases now available, and new more refined ones are under construction. A further division of the anatomical 'experimental' resources can be considered along the spatial scale axis. At the cellular and mesoscopic level there are numerous resources devoted to gene expression and strain anatomical differences. A smaller set of resources provide subcellular information. Integration between the resources is very much at its infancy.

Recommendations

- With regards to the current resources, desirable is an environment that presents a unified framework to search available resources and view retrieved data. Equally important are expanding the datasets available both by mining the literature and by creating new datasets.
- At the mesoscopic level, there is a need for a larger number of strains, larger representatives of each strain, and more developmental time points to support morphometric analysis. At the cellular level, gene expression has dominated efforts in the field. Sorely lacking is proteomic and detailed connectivity data.
- The data is not the only thing missing. While the spatial registration tools available yield adequate technology for data mining at the mesoscopic scale, informatics solutions to mine cellular level data, where one-to-one mapping is absent, are needed.
- Efforts must be made to integrate existing data resources into this atlasing framework, methods and tools for how to do this are covered in other sections of this analysis. To facilitate integration of new data, it is clear that standardization and data collection best practices need to be developed and circulated for use and improvement.
- Finally, data repositories that are linked to the digital atlasing framework must be provided that allow small groups to share their data without having to build and maintain their own databases.

D.2 Protocols and Tools for Preprocessing

Summary

Pre-processing has a lot to do with the how and why of the researcher, which means there's a great deal of variability in this area. There doesn't seem to be a standardized protocol to create an atlas; it is difficult to compare the anatomy across five different MR reference atlases due to variations in contrast and resolution.

Recommendations

- Suggest creating a starting point for standardized protocol for acquisition or calibration standard, or a minimum criteria e.g. examine data across different scanners, take a look at distortions, etc.
- Orientation of the data and transformations are defined via registration of the data. Images need to be tied to conventional histology
- We can develop a set of best practices for data collection protocols that would aid to bring that data type into atlas space
- Start with the basic information and look for minimum information across specific data types

All our recommendations must both allow registration and move us towards the ability to analyze across data experiments.

D.3 Spatial Registration

Summary

Registration of image data to a common reference spatial coordinate system is necessary to compare between "experiments" and to enable fusion of the information content to facilitate potential data mining and discovery. Ideally, once correspondence has been achieved one can transfer information from the reference space to the "experimental" image coordinates and vice versa.

Image registration for medical and clinical applications is a very active research area spanning a large body of literature and with the availability of numerous free and commercial tools ranging from simple command line tools (e.g. AIR) to full GUI workflow packages (e.g., Amira). Many of these tools can potentially be reconfigured to support the task of registering rodent brain image data to the atlas portal.

In context of rodent neuroinformatics, there are many challenges to obtaining registration of sufficient accuracy. First, there is inherent nonlinear variation between brains of different animals. Many biological experiments require the removal of the brain from the skull before imaging, resulting in further morphological differences. Division into histological sections may cause distortion and sometimes tears and folding of the tissue. Further, staining and other treatment could impart spurious smears and spot artifacts.

In general image registration techniques/tools can be classified by the:

- Registration metric
- Spatial transformation model and optimization algorithm used

The *registration metric* is a quantitative measure of the "goodness-of-fit." A measure can be landmark (point, corners, curves etc) based with the goal of minimizing the "distance" between corresponding landmarks. Landmark based schemes necessarily relies on robust (either manual or algorithmic) identification of correspondence. Alternatively, the registration criteria can be based solely on the image intensities (scalar or vector values at each voxel/pixel). Mean square difference and correlation are examples of simple intensity based measures. However, these simple measures are only effective for intra-modality problems where corresponding image areas have approximately the same intensities.

In the literature, mutual information (MI) (Amira, ITK, Analyze, 3DSlicer) has been described to solve various multi-modality problems. MI is an information theoretic entity that measures how much information is gained by one random variable (RV) (intensity in one image) by the knowledge of another random variable (intensity of another image). MI is suited to multi-modality problems since the actual form of the dependency between the two RVs does not have to be specified. However, MI should not be considered a panacea to all problems as the maximizing solution to MI may not also be the same as a true correspondence.

The *spatial transformation* model defines the allowable motion for the image to bring it into correspondence to a second image. Essentially, the model defines the parameter space for which we numerically *optimize* the registration metric using some search type of search algorithm (e.g., steepest descent, quasi-Newton, genetic algorithms). Models can range from simple rigid (translation, rotation) to linear (isotropic scaling, anisotropic scaling, shearing) to free-form deformation (b-spline or polynomial basis functions) motion.

Also of note are biomechanical-based solutions in which two images brought into alignment are modeled as a physical process resembling the stretching of an elastic material and thus leveraging mechanical analysis techniques such as finite element methods (FEM).

A classification table of some of the registration tools surveyed is provided below.

Desired State

Choosing the right registration technique/tool to use is highly dependent on the problem to be solved (single vs. multiple modalities; large capture range vs. requirement of close initialization). Practically, it is unlikely any predetermined set of registration "tools" can effectively cover the entire spectrum of problems. It is thus desirable to (a) allow registration savvy researchers to use and develop their own methodologies and (b) support the dissemination of these new techniques back into the public domain.

To support those researchers that are not interested or do not have the resources, the availability of a suite of easy-to-use reconfigurable tools is also desired. This tools framework could also be used in the registration workflow design stage where a researcher can mix and match various strategies (e.g. metric, transforms) to explore their particular problem space to identify problem areas for further development.

Another practical note is that naïve application of registration tools most likely will not work. Robust strategies typically adopt a "coarse to fine" approach where registration is first performed in a low resolution global sense (for example, using smoothed down-sampled images and a simple metric and transform model). The result from one pass is then used as the initialization for the second higher resolution pass (image content and/or transform and metric) and so on.

A highly desirable outcome of this project is to unify registration tools into a well thought out "open source" toolkit with the goal of extensibility by the user community as new techniques appear in the field. The "generic registration frame-

Table D.1: Survey of existing registration tools

	AIR	Amira	SPM	ITK
Landmark-based		Corresponding points		Supported
Intensity-based	Yes	Yes	Yes	Yes
Metric	Ratio image uniformity, partitioned image uniformity	Mean square difference, correlation, mutual information, normalized mutual information	Summed square difference in a Bayesian framework	Numerous. User can develop and incorporate more metric
Transform	Rigid, polynomial basis function	Linear (rigid, isoscale, anisoscle and shear)	Linear, low frequency basis function, vector field	Numerous. User can develop and incorporate more transforms

work" of the ITK toolkit is an example of such an extensible paradigm that mix and match different registration "components." Since each generic "component" has a standard API, new components can be added that can immediately make use of all the existing classes.

Recommendations

- While registration techniques can be used to bring image data into some common spatial coordinates, more effort is needed to define what information we want to extract from the image and how that information is to be queried and analyzed. For example, while it is possible to co-register high resolution 2D ISH images to a 3D framework – underlying desirable knowledge is the location of "expressing" cells within the spatial coordinate of a brain. Knowing what the downstream requirements will help to define standard API and other interoperability needs of the registration tools.
- The first step is to define a strong standard for defining coordinate system. This is essential as the overarching goal is to unify data from many different sources with different dimensionality (e.g., 2D, 3D, time series), resolution and orientation.
- The next step is to define a standard API to encapsulate the concept of spatial transformation. This is the key to allow integration/transfer of information content between the standard space and the image data. This "transform" API has to be generic enough to cover the spectrum from simple rigid transform to vector field to mesh models. Standardizing the "transform" API rather than "registration tools" will allow the community to use a variety of existing tools by adding a conversion layer.
- Another important task is to represent missing and sampled data. Biological data is typically obtained with high in-plane resolution but very low z-axis sampling and only cover a portion of the brain, for example, data focused on particular structures or a single hemisphere. Also, tissue damage can occur, resulting in "missing" data for some regions. This issue is important for the overall aim of using the data as a unified whole for data mining and discovery.
- Develop or support existing open source efforts to unify registration tools development in the community.

D.4 Database Upload Tools

Summary

A limited variety of different rodent-related neuroscience upload tools were examined for this analysis, mostly because it was difficult to identify many sites that offer public data uploads. It is not surprising that these interfaces are highly specialized for the database (DB) to which it is tied. These upload interface tools are commonly in the form of a web-interface, but some sources also offer the ability to upload data (usually to help with bulk uploads or frequent uploaders) and metadata in a standardized file format (e.g. Excel, csv, XML), and one offers a federated database system tied to the upload tools (a human DB used as a model for a rodent DB). There seems to be a wide range of ease-of-use and documentation for these tools. Finally, two additional issues emerge: 1) it is much more difficult to examine data from sites that offer loosely structured uploads, and 2) most upload sites request heir own proprietary standard format for upload.

Desired State

For upload tools, it would be ideal if there were easy to use resources for both data contributing biologist researchers as well as for people building new upload tools.

For Researchers

First of all, upload of data and associated metadata tools should be accessible and extremely easy. This would require easy to use interfaces for web-uploads, access to data processing workflows to ensure data are in the appropriate format, and tools that easily create bulk upload information in a standardized input format (e.g., Excel, csv, XML). These tools would access suggested ontologies/ terminologies for the researcher to appropriately and easily tag their data and describe their experiment before they upload it.

Also, if there was consistency between different database uploads, it is more likely that researchers would upload data more frequently and of different types. If experiments were related, but of different data types and potentially uploaded to different databases, it would be good for the researcher to only have to describe the experiment once and be able to upload it to different databases with minimal effort. Also, once uploaded, the data from a related experiment could be linked across the two different databases.

For Tool Developers:

Resources should be provided that would both make their job easier and also include standards for file formats, a set of tool packages that could be used with their own databases, and ontology tools and services that could be accessed for their needs. Ideally, these would also be designed in a manner that facilitates the interoperability of uploading data across different sources.

Recommendations

We recommend creating or supporting the resources listed below.

Standards:

- Standardized input file formats for bulk upload (e.g., Excel, csv, XML)
- Suggested ontologies/terminologies sources for a data type and services that allow access to these sources
- Suggested ontologies/terminologies sources for describing a subject and experiment and services that allow access to these sources.

Specialized Tool Modules:

- Web-based data upload modules that can be easily modified and applied to upload to a DB, along with easily modifiable documentation
- A tool package that will easily create the standardized input file formats for bulk file upload
- Examples of how to make use of and access the suggested ontologies/terminologies
- Usability suggestions/guides

D.5 Query Tools

Summary

What each tool exposes for querying depends on the purpose and research context in which the tools was developed, and more importantly on the managed data types and underlying data models. The data models are typically not explicit, so we cannot readily evaluate to what degree a particular tool exposes the underlying data structures. This is the main analysis difficulty.

Query tools we examined are quite diverse. The classes of tools include:

1. Predominantly gene expression/microarray repositories (e.g. ABA, GeneNetwork, Gensat, GEO, AGEA, ArrayExpress, NIH Microarray consortium, PhenoGen), with a range of form-based query interfaces, typically supporting search by genes and anatomic structure, as well as by subjects and by some additional parameters specific to each source;
2. Atlas gateways to multiple resources, supporting both spatial and attribute search, and often resource registration, annotation and search across several repositories (e.g. MBAT, SmartAtlas). These tools attempt to follow a common coordinate system, to support spatial data integration.
3. Desktop visualization tools focused on specific data types and working over local data (e.g. FiberViewer, MRIWatcher)
4. Semantic query and bridging tools (NIF, Bonfire, Neuronames) which support advanced ontology management and semantic query. Combining the ontology engines with PubMed seems most promising.

The grouping is not absolute, as several of the tools are truly comprehensive and possess features from several groups (e.g., ABA).

Desired State

Ideally, each tool would expose one or several canonical data types, which would determine what types of queries are possible. Thus, it would be useful to systematize information models employed by different projects, and do the comparison for each model. There are at least 11 groups of data models relevant to atlasing, but this list is certainly not exhaustive:

1. Cell models
2. 2d images
3. 2d vector segmentations
4. 3d volumes/reconstructions
5. 4+D volumes
6. Time series
7. Surfaces
8. Gene expression
9. Connectivity
10. Phenotype/behavioral
11. Annotations

There are sufficient examples of querying gene expression data. There is not much progress with other types of data, e.g., querying 2D and 3D images typically relies on pre-built correlations; there are just a couple tools that support graphical spatial querying, or integrated spatial-attribute queries.

This is a fairly artificial and abrupt grouping because in reality the data types may reflect different phases in data processing workflows: from 2D images, image mosaics and image stacks to 3D reconstructions, segmentations, annotations, etc. However, it is still useful to differentiate by data types because they would support different operations, and expose different properties for querying.

- Ideally, tools would provide web service APIs in addition to web interfaces, so that queries could be done programmatically. ABA, Smart Atlas, ArrtayExpress and GeneNetwork have published web service APIs. The APIs are at very different levels of maturity, and may provide standard-compliant XML output, or just formatted text. It would be useful to analyze commonalities between the APIs, and whether a common set of signatures can be developed within each data type.
- Query tools should adhere to common (or at least explicit) spatial, temporal and semantic frameworks, to ensure that query results from different sources can be combined. Significant effort has been made in establishing semantic consistency (with tools such as Bonfire, NIF, Neuronames). A common spatial framework is more difficult to attain (even with efforts of SmartAtlas and MBAT teams), hence the field is lacking truly graphical and coordinate-based querying and query processing.
- Ideally, query tools shall be capable of preserving user priorities and query history. Currently, only GEO does this and it is also planned in MBAT. Also, some tools keep selection state for the duration of the session (e.g., SmartAtlas and PubMed).
- Query tools should allow querying in multiple languages.

This places additional burden on query pre-processing services and lexical and ontology crosswalks.

- Embedding query interfaces in common data handling and visualization tools (e.g., Photoshop) would lead to quicker adoption by the community. The applications should rely on a uniform set of data retrieval and data discovery services across different datasets within each data type.
- Even common notions are somewhat problematic across different projects (e.g., the notion of "annotation" as adopted by Gensat, MBAT, and Smart Atlas, is not the same). A special effort within the atlasing community is required to establish semantic agreements.
- Querying across multiple databases implies standardized data discovery and access interfaces, single sign-on authentication, and establishment of metadata catalogs (for each data type) which would cache metadata from each source, to enable fast discovery queries.

Recommendations

- INCF shall discuss and determine the architecture strategy for the atlasing information system (clearinghouse/digital library, SOA, mediation, etc.), the types of data and metadata it shall contain, and the workflows it will support.
- Comparing information models used by different atlas query systems, and deriving common models for different data types, would be helpful.
- Establishing a common spatial, semantic and temporal framework, or at least making such frameworks explicit for each source, is important for developing a cross-source query engine.
- By establishing a standard for data query and retrieval services, and helping to ensure that key atlas data sources are accessible via such services, INCF will support a variety of client development efforts, including stand-alone clients as well as plug-ins for common commercial off-the-shelf (COTS) software.
- A common query execution environment should support search histories, user preferences, and cross-source authentication. INCF should emphasize internationalization of the query system (e.g., lexical translations, term mappings, etc.)

D.6 Annotation/Markup

Summary

This was a difficult area to examine since the big question still exists: What is annotation? For this group, this refers to tying space to labels. What is annotated? Brain regions, gene expression, other similar cell types and phenotype.

- Equivalent locations are often called by different names and use different hierarchies. This makes it difficult to cross-compare different atlases
- The amount of annotation varies widely
- Standard terminology is not standard

Recommendations

- A basic INCF standard annotation seems feasible. The community should be able to at least agree on the larger parts of the brain, while finer divisions will be more difficult
- Via spatial registration, it should be possible to view an ontology on top of an image and define to which atlas it is tied
- Need to determine what levels, as well as best practices for annotation are required for consumers, e.g. completely delineate an area within an atlas
- Technical details need to be worked out, e.g., how should people be tying information to the image; best practices: e.g., polygon, closed boundaries, defining voxels that can be converted
- Need a tool that allows a user to pull up different labels based on their needs

Most of these issues are shared with the Ontology program. The Digital Atlasing group will be more responsible for the technical components for allowing access to annotation tools and permitting people to access and create atlas annotations via the INCF Portal.

D.7 Analysis Tools

Summary

There are essentially three main classes of tools for the analysis of neuroinformatics data. These classes comprise tools designed specifically for the analysis of human data (e.g. SPM, FMRIB-FSL), those that that are organism independent and in fact designed for essentially any data type (e.g. Matlab, R/S+, ImageJ), and a number of model organism specific tools such as the small number of rodent specific applications (e.g MBAT, NeuroTerrain, NeuroBlast). Analysis of imaging data in the rodent in many ways represents a very small subset of the available analysis tools and scenarios in neuroinformatics analysis. Evidently, the majority of data analysis programs and methods have been developed for the analysis of human data, such as MRI, fMRI, DTI, PET, and related modalities. Most of these analysis scenarios have been motivated by clinical applications and are thus of secondary importance in the rodent. On the other hand, methods such as registration, signal detection, and classification are readily adaptable to rodent imaging and potentially applicable in that domain. Fundamental analysis suites such as Matlab, R/S+, C++ library methods, etc. can be configured to any model organism and thus low level libraries are available for interface and development. There are a handful of high quality methods and tools developed expressly for rodent image and data analysis and these should be the primary focus of a developing standard.

Desired State

A main concern of the present initiative is to enable access and interconnectivity of the methods specifically designed for rodent analysis, together with compatibility of at least a subset of the more generic tools. Whereas methods developed for human data are likely to be important, several may not apply and may be reserved for later efforts. The overarching goal of the present initiative is to insure that a unified perspective is attained for existing rodent methods. This includes the identification, access, and interoperability of these tools.

For Researchers:

From the research perspective, such an organization would imply access to high quality tools and interchange formats that facilitate comparative ease of analysis. A main consideration is that the existing rodent specific applications will remain isolated or independently developed in a manner that does not serve the larger research goals. A common and consistent analysis platform can be obtained if existing rodent specific tools are rendered interoperable.

For Tool Developers:

Rodent image and data analysis remains a small subset of research and clinical applications, and it is unlikely that the spectrum of applications will change. Tool builders might target work toward refining present applications and insuring interactivity of the existing methods. Secondarily, several of the methods available from human data analysis could be surveyed with the goal of use in the rodent analysis domain.

Recommendations

We recommend creating or supporting the following resources:

Standards:

- Examine the common modes of analysis used in the analysis of rodent data from the small set of domain specific tools.
- Insure that the existing formats and user requirements from existing rodent tools are consistent and interoperable.
- Promote the use of standard interchange formats that facilitate availability and access to basic tools.

Specialized Tool Modules:

- Existing rodent specialized tool modules should be enhanced and refined to enable ready access and interoperability.
- New tools should be developed within the context of extending or enhancing existing tools and made to easily interact with generic analysis programs or toolkits.
- Determine which methods from the wide spectrum of human analysis tools are applicable for the analysis of rodent data.

D.8 Integration Resources

Summary of Current State

An analysis of the current state of integration resources for atlases is to some degree an analysis of the current state of integration resources in general. In fact, for the purposes of these discussions we consider a simple three-layer model of integration resources. At the lowest layer we find domain independent integration resources and standards. The middle layer covers reusable integration resources in the biomedical domain, including those specifically dealing with atlas integration. At the top level we have applications that contain specifically developed integration components.

As always in such architectures, the objective is for each layer to reuse as much functionality of the lower layers as possible in order to maximise reuse. The level of domain and application-related knowledge required increases from lower to higher layers. The level of potential reuse of functionality increases from higher to lower layers.

On the one hand there is data integration which aims at linking several sets of data in a meaningful way, on the other hand one step further we find data fusion which aims at merging data in order to reduce the data. In terms of databases we think of data integration whereas on the lower level of images, i.e. binary data, data fusion is usually employed. In neuroinformatics/neuroimaging, images are merged in order to understand a particular configuration of the data visualized. The merging is done to a reference set which is most often an atlas. Data fusion on the image level is intended for analysis by a user and not for a data reduction.

Summary of State of General Integration Resources

Integration of computational resources, at the data as well as the tool level, has been a key topic in computing for decades. Integration may be achieved in a variety of forms, of which the most common today is hyperlinks (links between web pages). Others include database integration, integration through distributed workflows and the use of multi-agent systems (MAS). All of these have been used in the biomedical domain.

Hyperlinks in their most simple form are "hard-coded" (html) links from one part of a web page to another part either within the same web page or a different one. The links are purely syntactical, i.e. carry no semantic information about the link. Dynamic links, on the other hand, are created on demand, typically based on the content of some database. Semantic links associate information about the meaning of a link and may also be dynamic. For example, a term in a web page referring to a particular anatomical structure in mouse may be automatically linked to other web pages which are known (by the system) to also contain information about this structure; for semantic linking this is well elaborated in the CoHSE system [Bechhofer 2006].

Integration of databases has been an active research area for decades. Approaches vary depending on the level of autonomy each database retains. A commonly used approach is the database mediator. A mediator presents a single integrated view of the underlying databases, enabling applications and tools to treat the collection of integrated databases as a single resource. A mediator maps between the global integrated schema of the data and the individual schemas of the underlying databases and translates a query against the global schema into the appropriate sub-queries for the individual databases. In the context of atlas integration, the BIRN mediator is an example of an integration resource following this approach.

With the introduction of web and grid services as computational interfaces to distributed databases and tools, various workflow technologies have emerged. A workflow is simply the collection of a number of services where the output of one service provides the input to another service. A workflow thus presents a means of integrating several services in order to carry out some complex computational process. Systems have been developed for the creation and the execution (enactment) of such workflows. In the biomedical domain, two well-known examples are myGrid (www.mygrid.org.uk) and BioMoby (www.biomoby.org).

Originating from the Artificial Intelligence community is the notion of intelligent software agents. An agent may represent a particular resource, e.g. mouse brain atlas, a particular user or integration resources that communicate with the user and resource agents, resulting in so-called multi-agent systems (MAS). There are some similarities between agents and services, but agents are expected to have more built-in domain knowledge about how to solve complex tasks, making them more autonomous with respect to finding integration solutions. In contrast, the knowledge of how to combine services is explicitly built into a workflow by the workflow designer (typically a bioinformatician). For an overview of the use of MAS in the Life Sciences see [Burger 2007].

The more knowledge about how different resources relate to each other is built into an integration system, the more useful functionality it is able to provide. This is particularly true for the level of automation that is supported. In the case of links between web pages, the primary navigation and analysis task over the information brought together by these links remains with the end user. Database mediators, workflows and multi-agent systems, on the other hand, all aim at automating some of the integration work.

Much effort has been invested in the development of ontologies and tools for the engineering and use of such ontologies. A general discussion of this field is beyond the scope of this document.

Summary of State of Biomedical Integration Resources (including Atlases)

The most widely used form of integration of atlases and other biomedical resources is based on the link-model, i.e., hyperlinks from one resource to another, e.g., from the Edinburgh Mouse Atlas Gene Expression (EMAGE) database to the gene expression database (GXD) at the Jackson Laboratory. Typically, such links are dynamic, but not semantically marked up.

An integration mechanism for biological annotations with fairly wide uptake is the Distributed Annotation System (DAS). DAS is a client-server model where a single client gathers and displays information from several servers. The primary users of DAS are the genome bioinformatics community, but extensions (DAS/2) to the original protocol have been developed to extend its functionality. For further details on DAS, see www.biodas.org.

In terms of workflow integration, the two most widely used resources are myGrid and BioMoby. Although researchers can develop their own workflows, an increasing number of these are shared between scientists. The myExperiment project (www.myexperiment.org) currently holds over 300 biomedical workflows, typically for the myGrid system, which can be searched and downloaded. The Taverna system (taverna.sourceforge.net) is one of the more popular tools to browse and create workflows. It is noteworthy, however, that to date, neither the keyword 'atlas' nor the keyword 'anatomy' resulted in any search results in myExperiment.

The BIRN infrastructure is an example of a database-oriented integration resource, deploying the database mediator approach. This architecture is based on integrating ontologies while also deploying normalized coordinate spaces facilitating data fusion on the image level.

Many forms of integration depend heavily on the development, community acceptance and use of relevant standards for data and knowledge representation as well as communication protocols. In the biomedical field, we have seen major efforts in the development of so-called minimum information standards (MIS) as well as ontologies.

Recent developments of MIS include MIAME (**Minimum In**formation About a Microarray Experiment), MISFISHIE (Minimum Information Specification For In Situ Hybridization and Immunohistochemistry Experiments) and MINI (Minimum Information about a Neuroscience Investigation). While some of these have been widely adopted, e.g. MIAME, others are recent proposals, e.g. **MINI. The Minimum Information for Bio**logical and Biomedical Investigations (MIBBI) project is looking into collaborations across these various standards with a view to minimise duplication of work, and to encourage a level of interoperability across them. No atlas-specific MIS are available at this time.

Biomedical ontologies play an important role in the integration of resources. The Open Biomedical Ontologies (OBO)

web pages contain a collection of such ontologies. The OBO Foundry is a collaborative effort to promote the consistent use of a number of design principles for ontologies (www.obofoundry.org). The most widely used biomedical ontology is GO (Gene Ontology). With respect to atlases, the most relevant ontologies are those describing anatomy, which exist for human, e.g. GALEN and the Foundational Model of Anatomy (FMA), but also for many model organisms, including mouse, Drosophila, zebrafish, C.elegans, Xenopus and tick. Ontologies aimed at integrating multiple other anatomy ontologies include CARO, UBERON and MIAA. The latter, though entitled *Minimum Information About Anatomy,* isn't an MIS in the sense described above, but a simple vocabulary with mappings to other anatomy ontologies. In the context of BIRN, the BIRN Lex vocabulary has been developed.

MAS-based integration resources, although some have been developed, are not widely used as yet, as is the case for MAS in other domains. However, as the amount of domain knowledge modelled in integration resources grows, the paradigm of intelligent software agents is increasingly likely to replace, or at least complement, the notion of web and grid services as the primary computational component in distributed systems. The application of machine learning strategies that include the use of ontologies is emerging.

Desired State

Integration resources for atlases should provide reusable services and tools for a wide range of query and analysis functions that span multiple, distributed atlas resources as well as other resources relevant to information held in these atlases.

Specifically, atlas-oriented integration resources should facilitate the integration of spatio-temporal data which are based on heterogeneous spatial and temporal frameworks, thus facilitating the comparison of data from different sources.

Reflecting the different types of distributed computation, atlas-oriented standards (ontologies, protocols, APIs) should be available not just for link-based integration, but also for distributed databases querying, e.g. mediators, web and grid services and workflows, and eventually for distributed multi-agent systems.

Recommendations

Many of the issues raised and recommendations made in other categories, in particular the one for query tools, apply directly to this Integration Resources category as well. For example, the need for some common framework of spatio-temporal data is a recurring theme across the categories discussed. In general, the need for various forms of standardization has become clear. We will therefore limit the remainder of this discussion to points not yet raised elsewhere.

Assuming interoperability of atlas-oriented resources can be addressed and increasing numbers of such resources become publicly available, the need to describe these resources for easy discovery and reuse becomes more important. Furthermore, typical multi-atlas analysis scenarios will lead to common patterns of integration, which should be documented and made publicly available for scientists to share. In support of such developments, we recommend the following:

- Develop a systems ontology that can describe the semantics of the functions of atlas resources and their respective input and output parameters.
- A resource repository should be developed which holds descriptions for relevant atlas and atlas-related neuroinformatics resources.
- Standard integration problems and patterns for their solution should be recorded and made publicly available (including typical neuroinformatics workflows).

All of the above should take into consideration existing solutions; for example, those demonstrated by BioMoby, myGrid and BIRN.

References

Burger A (2007) Agent Technologies in the Life Sciences, Chapter in Semantic Web: Revolutionizing Knowledge and Discovery in the Life Sciences; Baker C and Cheung K-H, Eds. Springer, pp 341-52.

Bechhover S, et al. (2006) COHSE: Knowledge-Driven Hyperlinks the Semantic Web Challenge at the International Semantic Web Conference (ISWC).

Appendix E: Example Information Model

In a strawman example presented at the Digital Atlasing workshop, a simplified information model for gene expression (GE) data may include the following groups of web service methods:

- getSpecies, getSpeciesInfo,
- getSubjects, getSubjectInfo
- getGenes ({probeseries}), getGenes ({tissue}), getGeneInfo (gene)
- getProbes ({genes}), getProbeInfo (probe)
- getProbeTissues, getProbeTissueInfo (probe_tissue) [assuming a probe-tissue catalog being the center of a star schema]
- getGEValues ({probe_tissues})

In this example, all methods except the last will apply to the source's GE metadata catalog, which may be harvested into the central INCF registry as discussed in issue 5 of Overview of INCF Digital Atlasing Infrastructure, while the last method is a data access method executed against the source. If each GE source exposes such a web service API via a source wrapper as discussed in issue 2, then the system would be easier to scale and manage. Note that such information modeling is not expected to replace substantial modeling efforts in the community (which are, in case of gene expression/microarray data, reflected on MAGE/FUGE models), but rather extract "atlas profiles" of respective information models.

Appendix F: Key Resources for Early Employment of WHS

The following are key resources that we recommend adding as initial references and datasets in the initial deployment of Waxholm standardization.

Allen Brain Atlas (ABA)

A genome-wide, 3-dimensional map of gene expression in the adult mouse brain, the ABA reveals the expression patterns of approximately 20 000 genes throughout the adult mouse brain to the cellular level. Specialized tools have been developed to allow spatial access to the expression patterns in the dataset, including:

- *NeuroBlast* is a search tool to help identify genes with similar 3D spatial gene expression profiles. While searching for genes using conventional anatomic "search by region" is a natural approach to identify genes of interest, greater search power may be obtained by starting with a particular expression pattern and inquiring whether there exist other genes with a similar pattern of expression.
- A second project is the Anatomic Gene Expression Atlas (*AGEA*). AGEA characterizes the multi-scale spatial relationship in the mouse brain as derived from gene expression data without *a prior* knowledge of classical anatomy.

URL: www.brain-map.org

EMAGE/EMAP

The Edinburgh Mouse Atlas Project (EMAP) is a time-series of mouse-embryo volumetric models. The models provide a context-free spatial framework onto which structural interpretations and experimental data can be mapped. This enables collation, comparison, and query of complex spatial patterns with respect to each other and with respect to known or hypothesized structure. The atlas also includes a time-dependent anatomical ontology with mapping between the ontology and spatial models in the form of delineated anatomical regions or tissues. The models provide a natural, graphical context for browsing and visualizing complex data. The Edinburgh Mouse Atlas Gene-Expression Database (EMAGE) is one of the first applications of the EMAP framework and provides a spatially mapped gene-expression database with associated tools for data mapping, submission, and query.

URL: genex.hgu.mrc.ac.uk/Emage/database/emageIntro.html

GENSAT

The GENSAT project aims to map the expression of genes in the central nervous system of the mouse, using both in situ hybridization and transgenic mouse techniques. It is a collection of pictorial gene expression maps of the brain and spinal cord of the mouse. Using EGFP BAC-transgenics, this project provides the scientific community with tools to catalog, map, and electrophysiologically record from individual cells. The application of Cre recombinase technologies allows for cell-specific gene manipulation. The transgenic mice created by this project are available to the scientific community.

URL: www.gensat.org/index.html

Brain Architecture Management System (BAMS)

BAMS is an online resource for information about neural circuitry and connectivity. The rapidly expanding set of inference engines currently has 5 interrelated modules: Brain Parts (gray matter regions, major fiber tracts, and ventricles), Cell Types, Molecules, Connections (between regions and cell types), and Relations (between parts identified different neuroanatomical atlases).

URL: brancusi.usc.edu/bkms

NeuroTerrain Atlas

The objective of this group is to produce analytic tools, tissue resources, and genotypes essential for the systematic exploration of the complex genetics of mammalian brain architecture. The NeuroTerrain project focuses on development of atlases and brain normalization algorithms to support automated brain segmentation and image-based database query and works to incorporate data from the Mouse Brain Library (www.mbl.org). The library currently houses images of brain sections from over 2000 brains from different mice strains.

URL: www.neuroterrain.org

Cell Centered Database (CCDB) and SmartAtlas

The Cell Centered Database (CCDB) is a web accessible database for high resolution 2D, 3D and 4D data from light and electron microscopy, including correlated imaging. Techniques range from wide field mosaics taken with multiphoton microscopy to 3D reconstructions of cellular ultrastructure using electron tomography. The goals of the CCDB project include:

- Providing access for the biomedical community to primary and derived imaging 2D, 3D and 4D data from light and electron microscopy
- Developing advanced database capabilities for storing and mining complex cellular and sub-cellular imaging data
- Creating the necessary infrastructure for managing and sharing light and electron microscopic data securely within and between laboratories
- Developing tools and strategies for integrating data across scales and modalities
- Federating databases through the use of ontologies and shared spatial frameworks
- CCDB also holds images that are accessed via the SMART (Spatial Markup And Rendering Tool) atlas, a GIS-based

tool that supports coordinates and feature-based search ing over multiple databases, and it retrieves and overlays registered images and segmentations. This atlas is based on the Paxinos digital plates and because of licensing is not publicly available.

URL: ccdb.ucsd.edu/CCDBWebSite/index.html

GeneNetwork

The GeneNetwork consists of a set of linked resources for systems genetics. It was designed for the multiscale integration of networks of genes, transcripts, and traits such as toxicity, cancer susceptibility, and behavior. This open resource combines over 25 years of legacy data generated by hundreds of scientists with full genome sequences and deep transcriptome datasets. WebQTL is the leading GeneNetwork module, and has been optimized for online analysis of traits that are controlled by combinations of allelic variants and environmental factors. WebQTL exploits several genetic reference populations (GRP) of mouse (BXD, LXS, etc.) and rat (HXB).

URL: www.genenetwork.org/home.html

Mouse BIRN MRI Atlases (Caltech, Duke, UCLA)

A number of 3D MR mouse atlases are offered by this group, consisting of a delineated image volume and associated labels. All atlases can be viewed in MBAT or SHIVA, which allows navigation through all three planes as well as an arbitrary plane. In addition, BrainGraph is linked to the labels and allows visualization of the structure hierarchy and can be used to navigate through atlas. The current atlases offered in this format include:

- Mouse 3D MR, minimum deformation atlas from 11 female mice
- Mouse 3D Atlas based on magnetic resonance microscopy (MRM) images
- Mouse 3D MR Neonatal (P0) atlas and associated 3D Nissl Neonatal (P0) atlas

In addition, this group offers multiple datasets.

The Mouse BIRN group, which includes these three groups as well as the groups involved in NeuroTerrain, SmartAtlas, and GeneNetwork, has developed methods and tools for creating interoperability between MBAT, Smart Atlas, and NeuroTerrain. In addition, they have created the ability to access and display some of the gene expression information held in GeneNetwork from MBAT, and other types of information held in BAMS and BonFire (a BIRN ontology tool).

URL: www.nbirn.net/bdr/mouse_atlas_dti/index.shtm

High Resolution Mouse Brain Atlas (Sidman), BIRN CIVM Developmental Atlas

- An atlas of the developing mouse from E10.5 through E19.5 with PND 0, 2, 4, 8, 16, and 32 days using actively stained specimens at 19 micron isotropic resolution
- More than 200 labeled structures
- Labeled data is supplied in MBAT format or is directly downloadable as 3D-TIFF

URL: www.civm.duhs.duke.edu/devatlas/index.html

BIRN Mori Mouse DTI Volume Atlas 1.0

This resource provides an image atlas of developing mouse brains from embryonic and adult mice. The images were acquired using 3-dimensional diffusion tensor magnetic resonance microimaging techniques. The atlas includes viewing software and images from E14, E15, E16, E17, E18 and adult mouse brains.

URL: www.nbirn.net/bdr/mouse_atlas_dti/index.shtm

High Resolution Mouse Brain Atlas (Sidman)

The goals of the High Resolution Brain Atlas project from Harvard are:

- Create a full 2D digital atlas of a two-month-old female C57BL/6J mouse brain at 10 micron voxel resolution. The atlas is based on a complete set of coronal sections stained alternately for cells.
- Create a 3D voxel atlas of the C57BL/6J mouse brain from the present 2D section images. The segmented volumes will then be connected via an intuitively easy user interface to the informatics database of the High Resolution Brain Atlas.

URL: www.hms.harvard.edu/research/brain/intro.html

EuroExpress/GenePaint

EURExpress is a multi-site European project with 12 partners and the shared goal of generating expression data for >20,000 mouse genes via RNA in situ hybridization (ISH) on sagittal sections from E14.5 wildtype C57Bl6 embryos. A transcriptome atlas of gene expression patterns will be created using an automated RNA in situ hybridization system (GenePaint), for which experimental procedures, data collection and display have been standardized. This project integrates existing European initiatives, such as mouse mutagenesis and phenotyping projects, which depend on detailed information of gene expression patterns. For a subset of genes, particularly those involved with human disease, expression data will be generated using human and murine tissue arrays.

URL: www.eurexpress.org/ee_old/technology/publish.html

Paxinos Reference Atlas

The Mouse Brain in Stereotaxic coordinates and the Rat Brain in Stereotaxic coordinates are standard classics in the field to which many studies refer. Efforts are underway at Elsevier in conjunction with the Allen Institute to translate the 2D structures in the mouse and rat atlases into 3D and to provide these models as public resources.

URL: www.elsevier.com

INCF Secretariat, Karolinska Institutet, Nobels väg 15A,
SE-171 77 Stockholm, Sweden
Tel: +46 8 524 87093 Fax: +46 8 524 87150
Email: info@incf.org www.incf.org

International Neuroinformatics
Coordinating Facility