

1st INCF Workshop

on

Mouse and Rat Brain Digital Atlasing Systems

February 13–14, 2007 - Stockholm, Sweden

incf International Neuroinformatics
Coordinating Facility

www.incf.org



1st INCF Workshop on Mouse and Rat Brain Digital Atlasing Systems

February 13–14, 2007

**International Neuroinformatics Coordinating Facility Secretariat
Stockholm, Sweden**

Authors

Jyl Boline, Michael Hawrylycz, and Robert W. Williams
Email: rwilliam@nb.utm.edu

Workshop Organizer

**Robert W. Williams, University of Tennessee Health Science Center,
Memphis, Tennessee**

Workshop Participants

Jan G. Bjaalie, INCF Secretariat, Stockholm, Sweden
Jyl Boline, Laboratory of Neuro Imaging, UCLA School of Medicine, Los Angeles, USA *Reporteur
Gregor Eichele, Max Planck Institute of Biophysical Chemistry, Göttingen, Germany
Francoise Gofflot, Institut Clinique de la Souris (ICS), Strasbourg, France
Teiichi Furuichi, RIKEN Brain Science Institute, Wako, Japan
Michael Hawrylycz, Allen Institute for Brain Science, Seattle, USA
Andreas Hess, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
G. Allan Johnson, Duke University, Durham, USA
Jonathan Nissanov, Drexel University, Philadelphia, USA
George Paxinos, University of New South Wales, Randwick, Australia
Charles Watson, Curtin University of Technology, Bentley, Australia
Ilya Zaslavsky, University of California, San Diego, USA

Invited but unable to attend

Johan Auwerx, Strasbourg, France
Mihail Bota, University of Southern California, Los Angeles, USA
Maryann Martone, National Center for Microscopy and Imaging Research,
University of California, San Diego, USA
Larry Swanson, University of Southern California, Los Angeles, USA
Arthur Toga, Laboratory of Neuro Imaging, UCLA School of Medicine, Los Angeles, USA

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

Contents

1	Executive Summary	5
2	Introduction	6
3	Concepts	7
4	Workshop Discussions	8
5	Recommendations	9
5.1	Infrastructure For Data Sharing	10
5.2	Large Data Collection Project	13
5.3	Other Potential Proposals	15
6	International Collaborative Atlasing Projects Fostered by this Workshop	15
7	Future Relevant Workshops	16
	Appendix: Workshop program	17
	References	18



1. Executive Summary

The recommendations of this workshop group center around the mission of INCF “to contribute to the development of scalable, portable, and extensible digital applications that can be used by neuroscience laboratories to further knowledge of the human brain and related diseases.” Our expectation is that in the near future, research will involve intense interactions among scientists and computer networks at the level of ideas and dynamic reusable data. Neuroscientists will place greater reliance on stable web-enabled knowledgebases and will move away from rigid legacy models involving static research summaries and one-way non-digital communication. Atlases and spatial indexes will play a fundamental role in this shifting research landscape and will evolve into critical resources for gathering, securing, analyzing, and communicating research.

In the next decade, the standard bearer of scientific progress—the primary research paper—is likely to transition from a sleek and static synopsis of results and conclusions to a more complete, dynamic, and re-computable encapsulation of data and interpretation. This transition is already evident in the fields of genomics and bioinformatics in which papers are often pointers to massive data sets, analytic tools, and other appendix material. It is likely that a typical 3rd-millennium neuroscience communication/publication will include accessible data sets with full metadata, far more complete details on experimental design, and an enriched discussion section including interactive content such as a forum for genuine discussion between the lead authors and a larger community of reviewers and commentators.

Web-accessible brain atlases and spatial indexes will undoubtedly be among the most important tools needed to transition gracefully from our static synopsis mode of publication to a data-rich, dynamic, multidimensional mode of scientific interaction. Atlases will be key query tools as they appeal to our preference for visual exploration of complex data sets. Thus, this is a crucial time for an international community of neuroscientists to begin to converge on a lingua franca for digital neuroscience atlases, less with the goal of enforcing conformity, than with the goal of building resources and tools that translate among existing and future atlasing systems and their terminologies. This kind of transnational and translational activity ideally matches the mission of the INCF in the domain of digital atlasing.

With these longer-term objectives in mind, the 1st INCF Mouse and Rat Atlasing Workshop was held in Stockholm in February 2007. Our first objective was to survey current activities and plans related to mouse and rat brain digital atlasing systems and to produce a broad international inventory of resources and ongoing efforts. The second objective was to review the range of techniques that are being used to build, normalize, segment, and label atlases and to examine what aspects of this technical work are redundant, compatible, and compliant across platforms. The third objective was to forge an international

network to foster increased collaboration and interoperability across national, linguistic, and funding barriers and to examine how to promote international collaborations in the future.

The final and most important objective of this workshop was to improve the impact of atlasing projects in the near term (5 years) while reducing costs and redundancy of these efforts. Funds spent in support of the INCF secretariat and the national nodes should ultimately be leveraged many times over. Atlasing efforts in each member country should be able to accomplish more as part of a coordinated INCF activity than they would in isolation. To attain this goal, the INCF secretariat and the supporting nodes strongly encourage integration, code reuse, and joint projects.

Researcher goals and platforms differ, and all participants at this workshop accept, and even encourage, a wide variety of approaches to brain atlasing. However, all participants also appreciate the enormous benefits that may be gained by combining and integrating across diverse resources.

Consensus and Agreement

- There is a growing need for international digital atlas based neuroinformatics infrastructure as an organizing entity for data sharing and relating previously disparate knowledge bases into a linked system.
- The INCF is in a unique position to help coordinate the creation of standards and tools to harvest and maintain data and metadata.
- The INCF should encourage open standards for protocol, tools, and even specific data sets and atlases. There are now several “open access” images of the mouse brain (Allan Brain Atlas, Franklin and Watson, and NeuroTerrain). However, corresponding “open access” 3D segmentation data sets are still largely unavailable. Efforts by groups such as those at Drexel (Nissanov and colleagues) and UCLA (Toga and colleagues) are leading the way to open access to 3D atlases.
- There is a need for much more accessible data on connections among CNS regions and neurons incorporated into atlases to enrich functional network analysis (preferably open data sets of connections).
- Infrastructure and tools developed should be open, generally available, and easy to use.
- Any infrastructure work in this area depends heavily on database persistence and sustainability and ontologies to describe relevant data. Any work in this area will be influenced by the outcome of these upcoming INCF workshops.

2. Introduction

Publications are currently the primary way in which researchers share their findings, data, and metadata with the scientific community. In the past, the data from these experiments have rarely been used again and other researchers have often found it difficult to duplicate an experiment. More recently, there has been a push to share the original data, metadata, and even the processed and analyzed data with the rest of the scientific community. The added value and multiple applications of this are obvious: evaluation of the initial experiment by other investigators, the use of already collected data in other experiments, and new meta-analyses. This gives researchers the ability to examine issues in a way no single laboratory would have the resources for and to potentially answer questions or even pose new questions previously impossible.

Sharing data on a large scale with minimal interactions between the data provider and the data consumer creates the challenge of how to do this in a usable and meaningful manner. While it may be relatively easy to reduce the barrier to sharing, it is difficult to do so to a degree that still facilitates the process of data rediscovery and use for other purposes. Data repositories following this model have often turned into effective graveyards for data. Therefore, data discovery and data sharing must have associated information that will allow a scientist to concentrate on creating interesting hypothesis, not on the details of gathering the data and tracking down the methods of data collection. This may make it more difficult for the data provider, but it is better long-term insurance that their data may be used again for other purposes.

Much value can be derived from a data sharing system that finds and integrates data of many different types and from different sources. For this reason, digital atlases have been identified as potential neuroinformatics frameworks, as they act as a “map” one may use to traverse the brain and associated data of different scales, modalities, and sources. However, if we wish to have an atlas that acts as such a neuroinformatics hub, it must be more than a map. It must act as a gateway to large distributed databases of images, volumes, segmentations, and other types of spatially-registered data. These databases must

be connected through spatial data registries and services, as well as standard APIs and vocabularies to make them all work together. Contributing data within this framework requires tools for registration, image segmentation, spatial selection, and analysis. Finally, accessing this system requires viewers and annotation tools with authenticated access. In summary, there must be a full and complex infrastructure built behind the atlas as well as intuitive interfaces for interacting with them.

Several investigators have worked on various aspects of this issue, but it is not usually in an individual biological researchers’ best interest to pursue creating an infrastructure that is extensible to others at the neglect of their own unique research. As of yet, there remains a gap in all the potential resources and a full extensible sharing infrastructure in this format. Domain experts in this field would gladly participate in creating such an infrastructure, but most do not have the technical resources to build it on their own. What is needed to create such an infrastructure is an organizing body that can survey current practices, help generate standards, and aid in the technical construction of this sort of sharing infrastructure.

3. Concepts

To frame the discussion and establish the highest precision of intent it is first useful to establish concrete terminology.

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 as 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 Göting, <http://www.informatik.fernuni-hagen.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. (*IZ* 05/07)

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. (*IZ* 04/07, *MH* 6/07)

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. (*IZ 04/07, MH 6/07*)

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 04/07, MH 6/07*)

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*)

4. Workshop Discussions

Topics of the workshop included:

- An inventory and review of the production of digital rodent atlases (short presentations were given by all participants), their significance as models, and their diverse purposes including:
 - The display of gene expression and associated image data (Allen Brain Atlas, Cerebellar Development Transcriptome DataBase, GenePaint, Mouse BIRN Atlasing Toolkit-MBAT, MousePat, SmartAtlas)
 - Anatomic and genetic variation (Neuroterrain and the Mouse Brain Library)
 - Modeling neuronal circuits (Jan Bjaalie, Raphael Ritz and colleagues)
 - Comparative and functional analysis of behavior in rodents (Andreas Hess)
 - The advanced techniques needed to generate high resolution data as the backbone of these atlases (G. Allan Johnson, Gregor Eichele)
- A discussion of registration and mapping techniques used to align data to these atlases and the necessity for having atlases serve as a common frame of reference to which data is linked
- The applications of digital atlases as gateways to diverse data types, including gene and protein expression, patterns of neuronal connections, time-series records, population surveys, quantitative analyses, and species comparisons
- Access to stable and persistent databases with tools designed to easily upload and share multiple types of data as well as access and download this data
- The importance of defined and interoperable neuroanatomic vocabularies and higher order ontologies (George Paxinos, Charles Watson) to bridge the gap between different atlases, different species, different data modalities, scales, and as a consistent usable language between humans and machines
- The need for infrastructure, software tools, and technological expertise to facilitate data sharing

The participants of this workshop reviewed the current work in the area of atlas-based neuroinformatics and developed a set of recommendations:

- We reviewed the challenges of developing, applying, and translating consistent anatomical terminologies. Several examples were discussed including:
 - The challenge of defining basic structures, such as the amygdala, across atlases and data collection modalities. While the standard Paxinos and Swanson terminologies may be translated at a 3D stereotaxic level in adult rodents, how do we handle situations in which an experiment dictates that an area considered “amygdala” actually includes adjoining regions?
 - To what extent can homologies between CNS structures be assigned and defined beyond mouse and rat; mouse and human; mouse and chicken; mouse and zebrafish?
 - How best can terminologies (and atlases) handle the complexity of development?
- We reviewed experimental complexities and advances in many diverse areas of data collection:
 - New magnetic resonance histology methods
 - Function magnetic resonance methods applied to rodents
 - High throughput transcriptome and proteome methods in both adults and embryos
- We reviewed the values of atlases in examining experimental data and for tying different types of data backed by databases into a meaningful context including microscopy and electron microscopy data, other types of 3D volumes, surfaces, neural connections, blood vessels, gene expression levels experimental information
- We reviewed the power and problems of mapping data to digital atlases, an effect exacerbated by the lack of standardization

5. Recommendations

Through presentations, discussions, and follow up, this working group came up with two major recommendations for IN-CF’s role and participation toward achieving the goals outlined above. These are listed below and followed by a more complete description.

1. Encourage standardization and reproducibility in the development, maintenance and use of digital atlases by providing and maintaining a framework and/or infrastructure for digital atlas data sharing.
2. Serve in a guidance and organizational role for larger collaborative mouse and rat projects.

In addition, this group also made the following more general recommendations, which may be implemented in conjunction with or separate from the two above.

- Promoting International Recommendations and Standards (following examples of successful international standards bodies such as Open Geospatial Consortium <http://www.opengeospatial.org/>)
 - Make data sharing, standard API, and atlas software tool recommendations standard practice for code robustness, support for multiple interfaces, and better usability.
 - Best practices for managing and evolving the infrastructure in an extensible and scalable manner.
 - Gather and disseminate best practices learned across fields in a digestible format.
- Resources for Domain Experts:
 - Generate and garner financial resources for international nodes to build infrastructure and tools.
 - Aid in recruiting competent technical help for groups that wish to share their resources.
- Facilitate collaboration:
 - Facilitate communication between tool builders and users.
 - Identify groups for collaboration in areas of expanse.

- Identify groups for collaboration in other but related fields.
- Identify relevant technologies from other fields and parties that may aid in transferring the technology.
- Provide connectivity with people to help facilitate this collaboration.

5.1 Infrastructure for Data Sharing

The workshop committee recommends that the INCF support a vision of centralized standardization and infrastructure related to the development and maintenance of digital rat and mouse atlases. Rodent models are ideal for developing this infrastructure, as they are model organisms for human disorders, genetically defined, and may yield a wealth of diverse data types. Also, there are many existing atlases from which to generalize a standard. The ability to pull together these types of information both within and across species, greatly facilitates our ability to understand disease mechanisms and potential therapies. As the infrastructure for sharing is developed, it is expected that it will be extensible to other species and data types.

This proposed environment should facilitate data exchange, content discovery, and open interchange with existing data sets and formats. This recommendation would take the form of a potential set of standards for data sets and methods related to digital atlas including:

- Ontological and format proposals for a set of archival mouse and rat related neuroinformatics datasets with related metadata. These are the “canonical” datasets that provide a state of the art of rodent informatics imaging and methods.
- Description of the access methods for registration, mapping, and testing of the canonical datasets with a new test dataset.
- Use of a controlled open source environment with respect to archiving, access, and updating of the canonical datasets and methods.
- The ability to propose or upload other datasets for consideration in the canonical set.
- Reporting on performance and testing of the archival canonical datasets.

5.1.1 Identification of Canonical Datasets

The set of atlases used as the backbone of this project would likely include a set of digital atlases currently used by the community or any new high-quality datasets generated with this purpose in mind. These core datasets provide a new INCF standard for mapping and registration, whereas transforms enabling mapping of existing atlases into this new standard facilitate connectivity with legacy atlases and applications. The standard serves as the benchmark for future mapping, registration, and annotation efforts. Recommendations by this group for the initial set of canonical datasets and atlases:

Species/strain:

- Inbred strains of mice: primary = C57BL/6J, secondary strains include 129S1/SvImJ, DBA/2J or other of the 17 sequenced strains.
- Inbred rat strains including the sequenced Brown Norway BN strain and/or more widely used strains such as albino Sprague-Dawley.

Datasets: Provide the fundamental image datasets to which subsequent image registration and annotation would be referenced. Ideally this set should include:

- A multiple scan averaged high contrast 3D MRI volume (at least 30 micron resolution).
- A histology dataset, most likely a Nissl stain (at least 30 micron isotropic resolution, 1 micron in plane) that is reconstructed into a 3D image volume.
- Angiographic datasets would be collected in order to map the vascular system of the brain. This would include:
 - A MR angiographic dataset (at least 100 micron isotropic).
 - A μ CT corrosion cast to visualize vessels down to 10-20 micrometers in diameter.

The registered datasets would be used as the default standard for digital atlas applications and their comparison. Preferably, these would be from the same animal with an in-situ μ CT of the skull for the lambda and bregma references.

Age: Mouse within the range of 8-10 weeks; Rat within the range of 8-16 weeks approximately 350 g.

- Sex: 1 male and 1 female of each species/strain, dataset, and age.
- Fully delineated in a manner that enables the translation of known anatomy to significant structural depth (500+ structures). This should include registration mapping transforms from other widely used annotated atlases.

The inclusion of other types of canonical datasets over time might include:

- Other species and strains
- Different ages including embryonic stages
- Other annotations
- Marker genes
- Neural connections
- the spinal cord and the eye

Also the group showed a great desire to coordinate the assembly of a new annotated canonical atlas based on the same animal scanned from CT (for identification of Bregma), MRI, and histology.

There should be a curation process and methodology for bringing in any new canonical atlases established by the INCF that includes domain experts. Ideally, as infrastructure is built, tools would be offered that make it easier for people to comment on existing annotation in these atlases, or add their own.

5.1.2 Linking of Canonical Atlases

In order to use digital rodent atlases as the backbone for a neuroinformatics framework they need to be linked in a manner that will allow tools to navigate across different atlases and the data associated with them. The following are suggested methods for connecting these data:

Semantic mapping

- A consistent anatomic structure naming convention should be adopted, which includes both structure names and abbreviations (as GO has done for gene names). This has been a significant issue in the development of neuro-anatomic atlases to date.
- Structure names/Ontologies need to be mapped across atlases so it is possible to cross different labeling conventions.
- There needs to be a solid mapping between the rat and the mouse. It is likely this mapping will be semantic first, and spatial over time.
- It was suggested that as the group of canonical atlases are built to look for common features across atlases. These would be used as the core reference atlas, a set of unambiguously defined features and ontology that most scientists would agree with regardless of their preferences for how the brain should be delineated or named. These may be used to cross species.

Spatial mapping

- Registered to the same space, or explicitly defined coordinate framework (such as stereotaxic coordinates) which may also be used to define the framework for new data
- Spatially linked using an infrastructure that allows tools to translate from one atlas to another atlas

5.1.3 Web-accessible Platform for Access to the Canonical Atlases

The proposed platform would provide a framework and benchmark for the evaluation of new digital atlases, datasets, and methods. Offering maintained and refereed open source datasets and registration methods to these atlases would allow users to:

- Test and map novel mouse and rat datasets against canonical atlases of their choosing
- Improve existing and implement novel algorithms for mapping and translation between data sets

Sharing data within this framework requires data providers to upload their data using semantic and/or spatial mapping and may include whole brain, “chunks” of the brain, and several different data types. Thus, simple and accessible methods to enable this mapping must be offered. A specific architectural review of this proposed site and process should be performed to ensure optimal design and implementation strategies as well as its relationship to existing repositories/archives and datasets. The primary role of the INCF in this regard would be to set the policy and review recommendations for contributed material. Specific recommendations by this working group include:

- This platform would provide a test bed and archival storage system for registration methods against the canonical datasets.
- An architecture and site design review should be held for the housing of canonical datasets and mapping requirements.
- INCF would facilitate collection of registration methods and offer open access to their functionality by working with domain experts. Ideally, this will be offered through a simple web-accessible platform that helps step a user through the registration process in as close to an automatic process as is possible and practical. At the end of this process, the user must have the ability to automatically “publish” their dataset.
- Even with help, this group understands that registration is still a complex issue, and it may be necessary for INCF to assist people with registration (even if on a fee-based basis).
- There should be a standard of usability and a curation process and methodology for bringing in any new registration processes into the infrastructure that includes image registration domain experts.

5.1.4 Queryable Interfaces

It is vital to this project that data uploaded in this framework be stored in a manner that enables easy access, rapid content discovery, and comparison. It is advisable that the design and development of this interface be performed in concert with the atlas data upload aspect of the project.

- The databases (either new or existing-see section below) housing either the atlases or the newly uploaded data should supply data in a format of a common data model based call-service signature or syntax used for query call for that data type (API and/or web-services).
- Standard APIs will be needed for three levels of interoperability:
 - To retrieve different types of data from atlas servers
 - To query atlas metadata catalogs
 - To exchange information between atlas applications
- These common data model-based APIs and web-services should be defined by the experts storing and using that particular type of data (i.e. MAGE for microarray gene expression data, see references below). Since there isn’t a standard that has yet been adopted by that community, it is recommended that a global working group be created to develop one for that data type. This should include neuroscientists, database developers, query tool builders, and people with experience with developing this in other fields.
- Ideally the INCF would also provide a simple visual aid for web-query of this data. More specifics for the design of the query tools will also depend on the outcome of the survey of current software tools and practices.
- The issue of allowing users to create their own atlas delineations/annotations is complex and should be considered, although this would have to be implemented in a controlled manner. A primary goal of the infrastructure is to allow researchers flexibility while still operating within a set of standards that facilitate collaboration and sharing.

5.1.5 Existing Databases

There are a wide variety of related and supporting databases and information sources that are presently available and pertinent to this effort. Section 5.1.4 discusses how common data model based APIs and web-services will be key tools for operators of databases to may make their own resources available to the community query tools. Over the long-term, with the development of the appropriate resources to facilitate data sharing, different data modalities may be shared through this

environment such as electrophysiology, cell imaging, fMRI and disease phenotypes. To support these data sharing efforts, there must be tools, methods, and tutorials offered to the public that allow existing databases to link to this infrastructure.

The INCF should play a key role in this process. As these standards and methods are developed, INCF should make recommendations, tools, and tutorials available for how someone may be able to share their database with this infrastructure in a manner so that it is accessible for query and access.

Appropriate technical and domain people sponsored by the INCF may work with individual laboratories to implement the appropriate steps needed to share their data within this framework.

5.1.6 Software Tools

This proposed infrastructure requires a toolset that facilitates data upload, data registration, data query, and interoperable resources. As there are already many tools available in the community in these areas it is possible that tool providers will be interested in participating in this infrastructure, and that much existing infrastructure can be leveraged. An INCF approved study group would develop a plan for this framework based on a survey of what is currently being developed in this and other relevant fields. It is also essential that software developed for this infrastructure be open source and accessible for others to use and modify for their own needs.

5.1.7 Additional INCF Roles

While this is an ambitious project, this working group sees some very unique roles the INCF may play in facilitating the development of this infrastructure:

- Information gathering:
 - Survey neurobiologists who would use an infrastructure such as this about what data types should be the first areas of focus and what queries and integration scenarios they envision
 - Survey of available atlases and potential role they may fit in this architecture

- Based on these surveys, determine the canonical data types and atlases and possibly “canonical” databases or data sources
- Potential survey methods include examination of current resources and users per week, citation number, google scholar benchmarks, mailing list feedback, informal polling (asking experts and users etc. and propose alternatives)

- Oversee working groups involved in:

- Developing standards for data sharing in the areas outlined above
- Drawing up a technical specifications document that outlines how to implement this infrastructure and process
- Work with international groups to determine standards for facilitation of integration and implementation

- Develop database resources and ensure longevity and archival integrity of these resources

- Provide resources to aid in programming within the international nodes

5.2 Large Data Collection Project

Another role that the INCF is uniquely poised to play is as the parent organization for large data collection projects similar to HUGO’s role for the human genome project. This might include communication with funding sources, lobbying, and potentially housing of the database. The INCF would act as a critical guide and overseer for large scale high resource projects for which the value to neurobiology is high but for which there are concomitant risks.

As a project develops, the INCF could orchestrate planning groups that would develop concrete road maps and milestone checkpoints. As components of the project are identified, INCF would aid in recruitment of the expertise needed to fill the various roles in the project.

5.2.1 Example of a Representative Project

Recommendation for a Spatial Atlas of Gene Expression in the Rat Nervous System (as outlined by Gregor Eichele)

The rat is a widely used mammalian model system often used to study a broad spectrum of physiological, neurophysiologic and behavioral questions. However, much of this research is being adapted to the mouse, chiefly because of the relative ease by which genes can be manipulated in the mouse. Much effort goes into adapting a wide spectrum of assays from larger mammals to the mouse, but there are significant natural limitations to this such as tissue amounts, continuous monitoring of physiological parameters, and an intrinsic limitation of the mouse as a “smart” species. Instead of retooling the highly developed field of physiology, it seems more practical and less costly to re-determine the comparable genomic information in the rat. A first step has been completed with the recent completion of the rat genome sequencing project. Other techniques such as RNAi knock-down may make species such as the rat genetically tractable. Here we propose the production of a gene expression atlas of the rat nervous system.

The current project proposal is for a novel high resolution spatially mapped rat genome *in situ* hybridization atlas. The availability of key data modalities, such as the wealth of electrophysiological measurements in the rat, makes that organism the natural choice for a project that synergizes the connection of genetics, anatomy, and electrophysiology. In particular, the wealth of cell type specific electrophysiology in the rat such as the characteristics of ion channels will need to be coupled with gene expression data to untangle the computational complexity of the brain. Based on existing neurobiological resources, the genes of the rat genome can be prioritized so as to enable maximal coverage of significant expression patterns and cell types. Recent developments in neuroinformatics can be utilized to obtain maximal spatial resolution and image mapping accuracy.

5.2.2 Advantages of this Project

1. This is chiefly an engineering project that can be subdivided in clearly identifiable deliverables and milestones.
2. It is a highly interdisciplinary project covering experimental biology, high throughput lab methods, imaging and image processing, database technology and anatomical annotation.

3. Many of the goals of this project have been attained or addressed in a preliminary way in the mouse via the Allen Brain Atlas (ABA).

4. Several novel experimental and computational strategies not available at the time when the ABA was generated can be implemented.

5. There would be enormous benefit if this could become a multinational project similar to the human genome sequencing project.

6. Cost and time frame are predictable and finite.

7. The project uses in part existing resources but new ones could be added (e.g. a new interactive web database housing the expression patterns).

8. Genes can be prioritized by importance based on what is known from the ABA.

9. The price tag is in the range of 50 million €.

5.2.3 Implementation

In order to implement this project, planning groups representing the relevant components of the scientific community would need to be established. These groups would then draft the required concept papers and road maps. The INCF could potentially help in the organization and recruitment of appropriate parties. Example working groups include:

- Rat genomics experts to annotate genes and design templates for riboprobes
- Technology for 2nd generation tissue sectioning enabling improved section registration (e.g. high-throughput block face imaging)
- Logistics of tissue collection (stages of development, strain)
- Logistics large-scale template and probe generation
- Data collection (in situ hybridization and image capture, multiplexing for genes allowing detection of multiple transcripts per tissue section)

- Data transfer mechanisms, data processing and databases
- Computer-assisted annotation of expression by experts

5.2.4 Timeline

- 2007: Set up expert teams that will develop key points in sufficient detail (technically as well as financially) and generate by the end of 2007 a “road map.” Have identified potential contributors as well as funding sources.
- 2008/2009: Development phase for tools and adaptation of tools. This requires some pilot funding but is not yet a data production phase.
- 2010-2013: Data production phase, with establishment of database and annotation. This would be the more expensive part of project. Gene prioritization may reduce costs.

5.3 Other Potential Proposals

- Senescent C57BL/6J mouse brain examination of genes with age (NIA and Alzheimer’s) recommended by Dr. Robert Williams
- Christopher Reeve foundation requested a spinal cord atlas from Paxinos and Watson which could be integrated with ABA
- Pure anatomical connectivity-including quantitative information about these connections (including PNS input/output)
- High resolution rat MRI/histology
- Cytoarchitectural categorization of the volume of each structure and cell number within each structure to act as a normal resource for users to compare to their data (may be able to do with data already available)
- Compartmental models that include density of receptors etc.
- Ability to examine electrophysiological results and behavior.

6. International Collaborative Atlas Projects Fostered by this Workshop

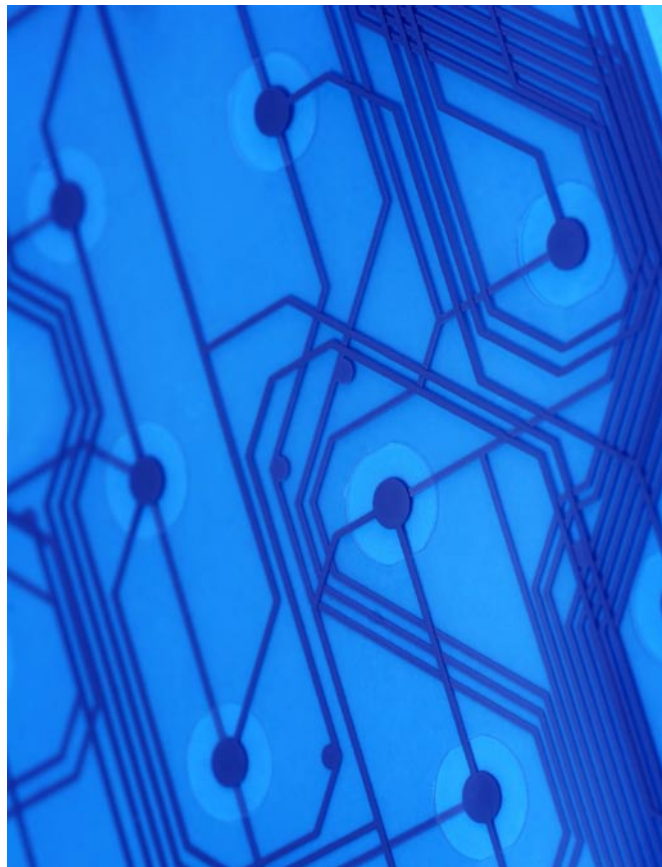
One of the goals of the INCF is to foster collaborations and help disseminate lessons learned, especially internationally. This workshop identified some potential international collaborations.

- The first collaboration discussed included Drs. G. Allan Johnson, Jonathan Nissanov, Charles Watson, and George Paxinos on a new rat atlas. This would involve a single rat that would be imaged by high resolution MR, followed by high resolution histology which would then be reconstructed into a volume. The Nissl volume would then be registered to the MR volume, and both used as the dataset for delineation. The highlight of this collaboration is to utilize the very high resolution and contrast methods of the Johnson laboratory.
- A second proposed collaboration was identified between Drs. Francoise Gofflot, Ilya Zaslavsky, and Jyl Boline. The goal is to link the MousePat database and their multiple gene expression efforts with the atlas and display tools of the Mouse Biomedical Informatics Research Network (BIRN) project.
- Finally, Dr. Gregor Eichle is exploring the possibility of creating a rat brain *in situ* atlas (see example collection project) that could incorporate techniques he helped develop that form part of the Allen Brain Atlas and GenePaint. Dr. Robert Williams suggested that one way to increase the scope and coverage of this effort would be to supplement *in situ* data for one strain and sex of rat with a microarray-based profiling study of 100 or more regions in both sexes of Brown Norway rats and in several other inbred rat strains, which is already funded by NIAAA. In addition, the BIRN group is developing tools that will facilitate incorporating and merging microarray and *in situ* data sets. This hybrid *in situ*-array method would permit rapid full-genome coverage at low spatial resolution (influenced by the array biopsy sample size and registration) but with the advantage of high spatial resolution *in situ* data for several thousands key CNS genes.

7. Future Relevant Workshops

This workshop focused on rat and mouse atlases; however, the development of this infrastructure is closely linked to other upcoming INCF workshops. The specifics for implementation of the plans outlined in this report will depend on decisions resulting from the outcome of these subsequent INCF workshops:

- Long term archiving of data/database persistence and sustainability – 1st INCF Workshop on Sustainability for Neuroscience Databases
- Ontology – 1st INCF Workshop of Neuroanatomical Nomenclature and Taxonomy (September 10-11, 2007)



Appendix: Workshop Program

February 13, 2007

12.00 – 13.00	Lunch and Introduction (Bjaalie and Williams)
13.00 – 18.00	Scientific presentations and discussions
Jan G Bjaalie	Database and spatial navigation systems for microscopic image data
Charles Watson	An ontologically based nomenclature for the mammalian nervous system
G Allan Johnson	Magnetic resonance histology of the mouse brain
Andreas Hess	Structural functional relationships in preclinical imaging, a domain for brain atlases?
Jonathan Nissanov	Spatial normalization of the mouse brain
Jyl Boline	Collaborative tool building for sharing and visualizing multimodal, multi-scale data
Ilya Zaslavsky	Integration of mouse brain atlas information across semantic, structural and application barriers
George Paxinos	Chemoarchitecture and rhombomeres as tools in constructing brain atlases
Francoise Gofflot	Systematic expression mapping of nuclear receptors in the adult mouse brain
Teiichi Furuichi	Spatio-temporal gene expression in the postnatal mouse brain
20.00 –	Dinner

February 14, 2007

09.00 – 12.00	Scientific presentations and discussions
Gregor Eichele	An outline of strategies for the production of transcriptome-scale 3D gene expression map of the rat brain
Michael Hawrylycz	Mapping and quantitation of high throughput in situ hybridization data
Robert W Williams	Variation in gene expression and morphometry in mouse CNS signal and noise
12.00 – 13.00	Lunch
13.00 – 17.00	Discussion

Each presentation was approximately 20 minutes, including questions. The sessions were primarily devoted to discussions and preparation of a draft report from the workshop.

References

- Bello M, et al. (2005) Hybrid Segmentation Framework for Tissue Images Containing Gene Expression Data., presented at MICCAI 2005.
- Blake JA et al. (2003) MGD: the Mouse Genome Database, *Nucleic Acids Research*, 31:193-195.
- Carson JP, Eichele G, and Chiu W (2005) A method for automated detection of gene expression required for the establishment of a digital transcriptome-wide gene expression atlas, *Journal of Microscopy-Oxford*, 217:275-281.
- Carson JP, et al. (2005) A Digital Atlas to Characterize the Mouse Brain Transcriptome, *PLoS Computational Biology*, 1:e41.
- Fitzpatrick JM, Hill DLG, and Maurer CR (2000) Image Registration, in *Handbook of Medical Imaging: Medical Image Processing and Analysis*, vol. 2, M. Sonka and J. M. Fitzpatrick, Eds.: SPIE Press, 447-514.
- Hajnal JV, Hill DLG, and Hawkes DJ (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.
- 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, in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2004: 7th International Conference*, Saint-Malo, France, September 26-29, 2004. Proceedings, Part I, vol. 3216, Lecture Notes in Computer Science, C. Barillot, D. R. Haynor, and P. Hellier, Eds., 192-199.
- Koslow S and Subramanian S (2005) *Databasing the Brain: From Data to Knowledge (Neuroinformatics)*, John Wiley and 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 Transactions on Medical Imaging*, 22:120.
- Ng L, Pathak S, Kuan L, Lau C, Dong HW, Sodt A, Dang C, Avants B, Yushkevich P, Gee J, Haynor D, Lein E, Jones A, Hawrylycz M (In Press) *Neuroinformatics for Genome-wide 3-D Gene Expression Mapping in the Mouse Brain*, *IEEE Transactions on Computational Biology and Bioinformatics*.
- Pluim JPW, Maintz JBA, and Viergever MA (2003) Mutual-information-based registration of medical images: a survey, *IEEE Transaction on Medical Imaging*, 22:986.
- Raya SP, and Udupa JK (1990) Shape-Based Interpolation Of Multidimensional Objects, *IEEE Transactions On Medical Imaging*, 9:32-42.
- Rex DE, Ma JQ, and Toga AW (2003) The LONI Pipeline Processing Environment, *Neuroimage*, 19:1033-48.
- Schwartz EL, et al. (1988) Applications Of Computer-Graphics And Image-Processing To 2D And 3D Modeling Of The Functional Architecture Of Visual-Cortex, *IEEE Computer Graphics And Applications*, 8:13-23.
- Sunkin S (2006) Towards Integration of Murine Spatial Resolution Expression Databases, *Trends in Genetics*, 22:211-217.
- Toga AW (2002) Neuroimage databases: The good, the bad and the ugly. *Nat Rev Neurosci*. 4:302-9.
- Toga AW and Mazziotta JC (1996) *Brain Mapping: The Methods*. San Diego: Academic Press.

Viola P, and Wells WM (1997) Alignment by maximization of mutual information, *International Journal of Computer Vision*, 24:137-154.

Visel A, Ahdidan J, and Eichele G (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, Thaller C, and Eichele G (2004) GenePaint.org: an atlas of gene expression patterns in the mouse embryo, *Nucleic Acids Res*, 32:D552-6.

Warren J, and 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 *Proc. 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, *Molecular Biotechnology*, 10:107-113.

Links of Interest

Scientific Computing and Imaging Institute, The SCIRun Problem Solving Environment, <http://software.sci.utah.edu/scirun.html>, 2004.

Microarray and Gene Expression Group, <http://www.mged.org/Workgroups/MAGE/mage.html>

www.incf.org

**INCF Secretariat
Karolinska Institutet
Nobels väg 15 A
SE-171 77 Stockholm
Sweden**

**Tel: +46 8 524 87 093
Fax: +46 8 524 87 150
E-mail: info@incf.se**



incf International Neuroinformatics
Coordinating Facility