

1st INCF Workshop

on

NeuroImaging Database Integration

August 30-31, 2007 - Stockholm, Sweden

incf International Neuroinformatics
Coordinating Facility

www.incf.org



1st INCF Workshop on NeuroImaging Database Integration

August 30-31, 2007

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

Authors

Lars Forsberg and Per Roland

Scientific Organizer

Per Roland, Karolinska Institutet, Stockholm, Sweden

Workshop Participants

Katrin Amunts, Forschungszentrum Jülich GmbH, Jülich, Germany

Jean-Pierre Changeux, Institut Pasteur, Paris, France

Rodney Douglas, UNI – ETH, Zurich, Switzerland

David C Van Essen, Washington University, St. Louis, USA

Lars Forsberg, Karolinska Institutet, Stockholm, Sweden (Rapporteur)

Jesper Fredriksson, KTH, Stockholm, Sweden

Albert H Gjedde, Aarhus University Hospital, Aarhus, Denmark

Kazuhisa Ichikawa, Kanazawa Institute of Technology, Kanazawa, Japan

David N Kennedy, Massachusetts Gen Hospital, Charlestown, USA

Torkel Klingberg, Karolinska University Hospital, Stockholm, Sweden

Per Roland, Karolinska Institutet, Stockholm, Sweden

Ulla Ruotsalainen, Tampere University of Technology, Tampere, Finland

Ryoji Suzuki, Kanazawa Institute of Technology, Kanazawa, Japan

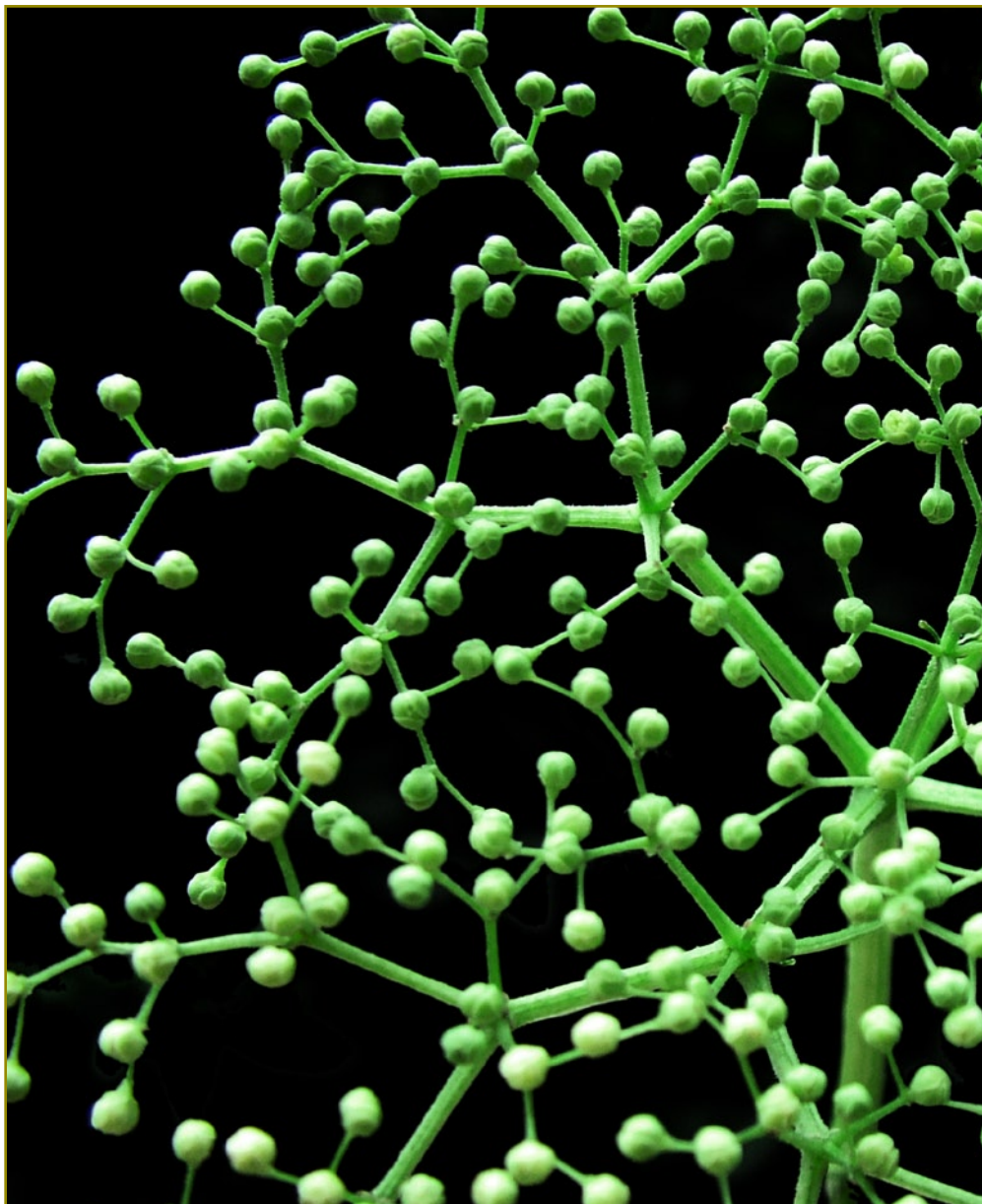
Jack Van Horn, University of California Los Angeles, Los Angeles, USA

Karl Zilles, Düsseldorf University, Düsseldorf, Germany

**Supported by the EU Special Support Action INCF, the INCF Central Fund and the Swedish
Foundation for Strategic Research**

Contents

1	Executive Summary	5
2	Introduction	6
3	Concepts	7
4	Federation of Databases	8
5	Recommendations	8
5.1	A Test Case—Data Sharing Between NeuroGenerator and fMRIDC	8
5.2	Future Development	8
	Appendix: Workshop program	10
	References	11



1. Executive Summary

The goal of this workshop was to map existing neuroimaging databases, particularly those containing primary data, and identify mechanisms that could facilitate integrated use of such databases, including interconnections between databases and data sharing.

1.1 Interconnections

The workshop group recommended that INCF should promote federation of databases by coordinating and integrating neuroimaging resources. Such a federation can be achieved by creating a portal from which different databases can interconnect to each other. The group recommended that interconnections between databases should be facilitated through the design and development of an INCF portal. This portal should give users access to all the databases through one interface, and also enable the users to go from one database to another and search for similar contents. Still to be determined is how to compare data between different databases through modeling.

1.2 Data sharing

A federation can also be achieved through data sharing between two or more databases, where the data from one database is integrated into another database. With this integration, data can be preprocessed and presented through the context of other databases. The group thought that INCF should promote data sharing and recommended that this should start with a test case: data sharing between the two fMRI databases fMRIDC and NeuroGenerator. The knowledge and tools developed in this project should be reused for future data-sharing projects.

With the knowledge gained from this project, INCF should help to advance the issue of describing neuroimaging data. This can be achieved by collaborating with major journals to develop guidelines and standards for reporting metadata related to experimental methods. Furthermore, INCF should encourage investigators to submit data and tools to a clearinghouse.

1.3 Future possibilities

The workshop group recommended that INCF should attempt to integrate databases of different modalities, including structural MRI, EEG, MEG, microstructural, (e.g., cytoarchitectonic) and receptor databases. Anyone should be able to access the databases, not only the member states of INCF. User login and password should not be required. The search features for the different databases should be documented on the portal with tutorials and examples.

2. Introduction

The very first functional neuroimaging database with online access on the Internet was created in 1994 [Fox, Lancaster, 1994]. Known as BrainMap, it currently contains 5400 experiments. Mosaic was still the most commonly used web browser at the time BrainMap was created and the size of hard drives was measured in MB rather than GB. Thus, given the infancy of the World Wide Web, the most realistic content to store in this database was the stereotaxic coordinates that were published in the scientific papers. Since most studies publish their results in the Talairach coordinate system, the results from different studies should be comparable. An MRI-based stereotaxic atlas from 250 subjects was published in 1992, to be followed by the MNI 305 template in 1993 [Evans et. al. 1992, Evans et. al. 1993]. However, the number of different stereotaxic standard brains continues to increase and although they should all be in the Talairach coordinate system, their coordinate systems are all significantly different from the Talairach system and from one another. Therefore, to compare studies with different standard brains, a transformation needs to be applied directly to the coordinates [Brett et. al. 2002; Lancaster et al., 2007] or through a surface-based atlas as an intermediate [Van Essen, 2005].

A functional ROI (region of interest) in a functional experiment may contain thousands of coordinates, which describe its extension and form in 3D space. Publications often only report the statistical peak coordinate of the ROI, together with its size, but the extension and form of the ROI are not published. Particularly when the ROI is actually multiple functional regions merged together in the thresholding of the statistical image, the peak coordinate from the most significant ROI is typically the only information reported. If a database also covered the form of the functional ROI, it would have to store the 3D image format which describes the full extension of the functional region.

In 1999, the European Computerised Human Brain Database (ECHBD) endeavored to store the findings from a functional study as 3D images, rather than just the peak coordinates [Roland et. al. 1999]. It also contained cytoarchitectonic measurements in the same stereotaxic space as the functional images. The database was unable to accept data from different standard templates, so the users were asked to change the format of their data into the ECHBD standard brain format before submitting it to the database. In practice, it was impossible to ensure that the data was correctly transformed, and the data would be difficult to compare. Additionally, different experiments were processed in different ways, with varying spatial filter sizes, different thresholding of statistical images, etc. One way to create a homogenous statistical database is thus to analyze the data through the same processing pipeline, using the same methods and the same standard brain. This would require the users to submit the raw data and enough experimental metadata to analyze the study in a standardized pipeline.

NeuroGenerator is the successor to ECHBD, and it has addressed the problem of heterogeneous datasets by collecting the raw data and processing it through a common processing workflow [Roland et al, 2001]. The result is a statistical database where the user can select which filter size and threshold level to work with. All the data has been transformed into the same standard brain using the same transformation method. The project was created in 2000 and the first database was sent out to users in 2003. Due to the size of the raw data, NeuroGenerator only allows people to access the statistical data. This access is via an open-source 3D visualization and query tool, available for Linux and Mac OS X. The database currently contains 67 studies from 649 subjects as well as cytoarchitectonic probabilistic maps for anatomical reference [Amunts and Zilles, 2006].

The fMRI Data Center (fMRIDC) is a public repository of peer-reviewed fMRI studies [Van Horn, et. al. 2001]. This repository was established in 1999. It currently contains 122 complete neuroimaging studies representing thousands of individual subjects, and hundreds of thousand of individual fMRI and MR structural volumes, as well as accompanying meta-data from published research articles. Users may request study data and it will be delivered to them on media and via digital download in cases where the overall size of the data is not prohibitive. The fMRIDC thereby allows the researcher not only to replicate the findings of the original study, but also to apply other methods and potentially obtain new or alternative findings, and even use the data in training and education.

The number of studies represented in NeuroGenerator and fMRIDC is much smaller than those in BrainMap, but the size of these databases is much larger because each study can be anywhere from 10-20 GB in size. BrainMap on the other hand, strives to record the reported statistical local maxima of activation from published articles along with the details of experimental design, thus facilitating meta-analysis across neuroimaging studies. The final derived statistical database from NeuroGenerator is a few 100 MBs in size. Depending on the meta-study, the researcher can benefit from all these different kinds of databases.

This report summarizes the INCF workshop on Neuroimaging Database Integration, whose goal is to map existing neuroimaging databases containing particularly primary data, e.g., NeuroGenerator and fMRIDC, and discuss the benefits and issues regarding data sharing across databases. One major topic was whether INCF should contribute to facilitating data sharing between fMRIDC and NeuroGenerator.

3. Concepts

In order to get a common ground for the following discussions, it is necessary to establish some basic concepts.

Integration

Integration of databases can either be to establish an interoperability between them or to share data between databases.

Interoperability

Interoperability refers to the ability of different database systems to work together (interoperate) using an established protocol for interaction.

Data sharing

Data sharing can refer either to sharing data through a clearinghouse or to copying data between databases.

Database

A database is a data repository stored in a structured way, with the possibility to query the repository for information.

Clearinghouse

A neuroimaging clearinghouse is a distribution center which collects and distributes tools and data.

Pipeline

A processing pipeline is an ordered set of tools forming a directed workflow used to process data.

4. Federation of Databases

The number of databases will continue to grow. Integrating neuroimaging databases will be made possible through a federation of databases. This integration will make the data more useful to researchers, allowing them, for example, to go from one database resource to another and see which other databases have common fields (e.g., anatomical regions). For this to become a reality there must be a portal through which researchers can access the different database resources and a way for them to easily jump between the databases. Data sharing is an alternative example, that involves preprocessing the content from one database and presenting it in a different way in another database.

The participants at the neuroimaging workshop agreed that data sharing is highly desirable, as it has the potential to accelerate research progress in neuroimaging, analogous to the dramatic advances that have occurred in genomics and other areas of bioinformatics. However, there were some concerns as to how to do this in practice while simultaneously maintaining the quality. More specifically, how does one compare data between different databases? There has to be a link between different observations and theories through modeling. In this process, it must be specified what kind of data to be compared and confront the collection of data through modeling. We have to understand what types of questions we are going to address through integration of databases. In the case of data sharing, what kind of data should be exchanged and should there be a minimum description of neuroimaging data to facilitate such data exchange? Should we be more precise as to what kind of data to compare and exchange from a theoretical point of view? How do we find the balance between simplicity of access (which would result in broad acceptance and usage by the community) and high quality?

5. Recommendations

5.1 A Test Case—Data Sharing Between NeuroGenerator and fMRIDC

The workshop gave rise to many questions, which is why the participants recommend that INCF attempt an example case study as a demonstration of how the scientific community might benefit from data sharing between two databases. It was felt among the workshop participants that the data sharing between NeuroGenerator and fMRIDC could serve as such a test case if justified by a well-defined proposal with realistic objectives. The role of INCF should be that of an honest broker, to facilitate fulfillment of the agreement. INCF can also provide technical support and resources, as long as the development can be reused in the future and if the process can be generic to other databases.

To get started, some example datasets would be exchanged in order to demonstrate the value of data sharing between NeuroGenerator and fMRIDC. During this process, quality control of the data is important so as to maintain the integrity of each of the databases.

There were some discussions about the metadata needed to get a precise insight of the conditions in an fMRI experiment, as well as the issue of minimum description of neuroimaging (for instance as an XML or RDF standard). For the data sharing between NeuroGenerator and fMRIDC, the decision was to not wait for an XML standard. The two databases already have their own description of fMRI studies. The challenge here is to establish a suitable exchange protocol between the two databases and to define suitable benchmarks as to what would constitute successful data sharing and integration in this test case.

5.2 Future Development

The proposed, as well as other, test cases, should help in identifying what a minimum description requires, so that it could be reused in future data sharing, including other databases such as Function BIRN. For future data collection, it is highly desirable to include a minimum description of neuroimaging data. INCF could help advance the issue of describing neuroimaging data and encourage people to submit data to an appropriate database. To get this started, it was suggested that INCF should coordinate with major journals to help in identifying

relevant requirements. One problem in the field today is the lack of a standard way to present the methods of a study in a scientific paper. This lack of standardization makes it difficult for researchers to compare different studies. A minimum format for describing neuroimaging and the statistical analyses will reduce uncertainties as to how the results were obtained, thus making it more likely that another study will be able to reproduce the results. This would also be a way to indicate to journals that the study meets appropriate standards. The Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC – nitrc.org) may be useful in this regard.

Demonstrated at the workshop was the LONI-pipeline, whose front end is a graphical workflow of different analysis modules. The processing backend can either be the same computer or a cluster of computers. The LONI-pipeline description can be used to present the statistical methods used in the publications. If data were submitted to a clearinghouse together with the necessary processing tools and a LONI-pipeline description, anyone could reproduce the results using the same data. By sharing computing resources, people might be more willing to use the LONI-pipeline system and finally submit the data in a standard format to a clearinghouse. This would increase the quality of the data.

In practical terms, it is not possible to define all the metadata needed to describe the conditions of an experiment in detail. The complexity of new experiments is increasing as new methods are being used to analyze the data. A minimum description is not intended to cover all methods and paradigms of all experiments in the way a fully-featured ontology would, but it should nevertheless attempt to describe a major fraction of a major subset of published studies (as opposed to all aspects of all studies). If journals supported a consistent minimum information framework as sanctioned by INCF, authors would have specific guidelines about what information should be included in their submitted scientific manuscripts, which would help to improve methods reporting. Furthermore, if this information were present, then text-mining approaches applied to the content of the published article would be enriched and better able to identify similarities between studies, etc. Minimal information provides several advantages for improving scientific communication and INCF should play a role in fostering the development of MIAMI-like lists of domain-specific metadata classes of these purposes.

The SumsDB database of structural and functional neuroimaging data was also demonstrated at the workshop. SumsDB contains a diverse set of surface-based and volume-based data from humans, monkeys and rodents. The human data includes over 14000 stereotaxic coordinates from more than 500 published studies, along with extensive experimental metadata. This data could potentially be federated with the aforementioned coordinate data in BrainMap, thereby capitalizing on the complementary visualization and analysis methods associated with the two databases. Additionally, the volume-based data in SumsDB could potentially be federated with the data and analysis resources of fMRIDC and NeuroGenerator.

Federation of databases can be described as either data sharing or interconnectivity between the databases. Comprehensive interconnectivity requires a portal from which one can access all the databases. One suggestion was to have a common search interface on the INCF website instead of choosing a specific database. It should also be possible to migrate between databases through the INCF portal or through other portals such as the Neuroscience Information Framework (NIF). To facilitate this process, it was recommended that databases should be evaluated by INCF and best practices should be established, possibly through collaboration between INCF, NIH, EU and Japan [Suzuki et al., 2007]. Federation of databases should eventually cover not only fMRI databases, but also structural MRI, EEG, MEG, and other modalities such as a microstructural (e.g., cytoarchitectonic) or receptor databases [Amunts and Zilles, 2006; Eickhoff et al., 2005; Zilles et al., 2002].

All the participants agreed that anyone should be able to access the databases, not only those from the member states of INCF. It should not even require a user login to access and query the databases. To help people understand how the databases could be used, there should be tutorials on the portal describing the search features.

All in all, a useful and successful database integration will benefit researchers greatly, creating better ease of use and improving the current standards in the field.

Appendix: Workshop Program

August 30:

09.00 - 09.30	Introduction and orientation (Bjaalie and Roland)
09.30 - 12.00	Scientific presentations and discussions
Karl Zilles	Brain maps for functional imaging: From histology to probabilistic maps
Katrin Amunts	Brain maps for functional imaging: From probabilistic maps to meta-analysis
Jean Pierre Changeux	Modeling access to consciousness and its consequences for brain imaging
Ryoji Suzuki	Neuroimaging study and platform in Japan: Overview
Kazuhiisa Ichikawa	Neuroimaging study and platform in Japan: Linguistic brain functions, integrative analysis, and NIMG-PF
12.00 - 13.00	Lunch
13.00 - 18.00	Scientific presentations and discussions
Jack van Horn	Is it time for a minimum data framework for neuroimaging reporting, exchange, and archiving?
Ulla Ruotsalainen	Databases of functional brain PET images
David C van Essen	Mining structural and functional neuroimaging data using the SumsDB Database and WebCaret visualization
Rodney Douglas	The challenge of EM image storage and analysis in large volume (m^3) reconstructions
Albert H Gjedde	Functionally integrative neuroscience: Expanding the frontiers of brain function
David Kennedy	The neuroimaging informatics tools and resource clearinghouse: A new knowledge environment for fMRI research
Lars Forsberg	Finding co-activation patterns with PCA: A meta-analysis study using the NeuroGenerator database
Jesper Fredriksson	Mining the NeuroGenerator database
Torkel Klingberg	Imaging of brain development and models of brain development
19.00 -	Dinner and further discussion

August 31:

09.00 - 12.00	Discussions and draft report
12.00 - 13.00	Lunch

Each presentation was scheduled for 20 minutes, including questions.

References

- Amunts, K., Zilles, K.: Atlases of the human brain: Tools for functional neuroimaging. In: Neuroanatomical tract-tracing 3-moleculesn Neurons, Systems (Zaborszky, L., Wouterlood, F.G., Lanciagi; J.L., eds), pp. 566-603 (2006)
- M. Brett and I. S. Johnsrude and A. M. Owen, "The problem of functional localization in the human brain", *Nature Reviews Neuroscience*, Vol 3, 243-249, March 2002.
- Eickhoff, S., Stephan, K.E., Mohleberg, H., Grefles, C., Fink, G.R., Amunts, K., Zilles, K.: A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* 25, 1235-1335 (2005)
- A. C. Evans and D. L. Collins and B. Milner, "An MRI-based stereotactic atlas from 250 young normal subjects", *Journal Soc. Neurosci. Abstr.* 18: 408, 1992
- A. C. Evans and D. L. Collins and S. R. Mills and E. D. Brown and R. L. Kelly and T. M. Peters, "3D statistical neuroanatomical models from 305 MRI volumes", *Proc. IEEE-Nuclear Science Symposium and Medical Imaging Conference*, 1813-1817, 1993.
- P. T. Fox and J. L. Lancaster, "Neuroscience on the net", *Science* 266, 994-996, 1994.
- Lancaster, J.L. and Tordesillas-Gutierrez D. and Martinez M. and Salinas F. and Evans A. and Zilles K. and Mazziotta,
- J.C. and Fox P.T. (2007) Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Hum Brain Mapp.* Jan 31; [Epub ahead of print].
- P. E. Roland and J. Fredriksson and P. Svensson and K. Amunts and C. Cavada and R. Hari and A. Cowey and F. Crivello and S. Geyer and G. Kostopoulos and B. Mazoyer and D. Poppelwell and A. Schleicher and T. Schormann and M. Seppa and H. Uylings and K. de Vos and K. Zilles, "ECHBD: A database for functional-structural and functional-functional relations in neuroimaging", Poster 128, Fifth International Conference on Functional Mapping of the Human Brain, 1999.
- P. E. Roland and G. Svensson and T. Lindeberg and T. Risch and P. Baumann and A. Dehmel and J. Fredriksson and H. Halldórsson and L. Forsberg and J. Young and K. Zilles, "A database generator for human brain imaging", *Trends in neurosciences.* 24:10, 562-565, Oct. 2001.
- Suzuki, R., Niki, K., Fujimaki, N., Masaki, S., Ichikawa, K. and Usui, S., "Neuro-Imaging Platform for Neuoinformatics", 14th International Conference on Neural Information Processing (ICONIP2007), Kitakyushu, 2007, TMB-2
- D. C. Van Essen, A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. *NeuroImage.* 28, 635-662, 2005.
- J. D. Van Horn and J. S. Grethe and P. Kostelec et. al, "The functional magnetic resonance imaging data center (fMRI-DC): the challenges and rewards of large-scale databasing of neuroimaging studies", *Philosophical Transactions of the Royal Society of London Series B Biological Sciences*, vol. 356, no. 1412, 1323-1339, 2001.
- Zilles, K., Schleicher, A., Palomero-Gallagher, N., Amunts, K.: Quantitative analysis of cyto- and receptoarchitecture of the human brain, pp573-602. In; *Brain Mapping: The Methods*, 2nd edition (A.W.Toga and J.C.Mazziotta, eds) Academic Press (2002)

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.org**



incf International Neuroinformatics
Coordinating Facility