NEUROIMAGING DATABASES AS A RESOURCE FOR SCIENTIFIC DISCOVERY

John Darrell Van Horn, John Wolfe, Autumn Agnoli, Jeffrey Woodward, Michael Schmitt, James Dobson, Sarene Schumacher, and Bennet Vance

The fMRI Data Center, Dartmouth College
Hanover, New Hampshire 03755

The field of neuroscience has an increasing need for access to primary research data in order to more thoroughly explore fundamental neural function beyond those examined in the original published article. For instance, functional magnetic resonance imaging (fMRI) studies of the human brain during the performance of cognitive tasks involve the collection of several gigabytes of image volume time course data as well as detailed meta-data concerning subject, experimental, and scanner protocols. Much of this data is unseen by anyone other than the original study authors but could be used by others to gain new insights into basic cognitive processes. We describe how several efforts have sought to archive the primary data from brain imaging studies and make them available to researchers in the community. We detail several aspects of neuroscientific data sharing that can help promote new inquiry. Essential in this process is the design of hierarchical frameworks for encapsulating fMRI study data for the purposes of extensible study organization that helps to encourage data sharing between collaborators or centralized data archives. In this chapter, we feature our own effort, the fMRI Data Center, as an example of large-scale archiving of fMRI study data from the peer-reviewed literature and how this is being used to explore data beyond the scope of the original article. Through such efforts, brain imaging has begun following the lead of the biological sciences by leveraging its accumulated data into new knowledge about fundamental brain processes.

INTERNATIONAL REVIEW OF NEUROBIOLOGY, VOL. 66
DOI: 10.1016/S0074-7742(05)66002-3

Copyright 2005, Elsevier Inc.
All rights reserved.
0074-7742/05 $35.00

I. Introduction

The ambitious attempts being made across the biological sciences to promote the sharing of data and to facilitate meta-analysis (Becker, 2001; Mavroudis and Jacobs, 2000; Mirnics, 2001; Nowinski *et al.*, 2002; Reidpath and Allotey, 2001; Richard and Williams, 2002) have now transformed the process of science in the digital age. Large-scale scientific databases and collaborating networks of researchers are being developed to enhance scientific interaction and promote the sharing of primary research information (Collins and Mansoura, 2001; Shepherd *et al.*, 1998). For example, scientific databases containing experimental data and results, such as the Protein Data Bank (Berman *et al.*, 2000), permit re-examination or comparison of data on 3D protein structures in order to test new hypotheses and enable data mining from across research centers to reveal trends that may give rise to new avenues of research. Much excitement has occurred in the neurosciences about large-scale databasing (Koslow, 2000) and the promise that they hold for understanding cellular properties (Martone *et al.*, 2002), neuronal models (Marenco *et al.*, 1999), and underlying biochemical pathways (Karp *et al.*, 2002).

In the 1990s, use of fMRI during performance of neurocognitive tasks rapidly emerged, surpassing the use of positron emission tomography (PET), as the fundamental tool for mapping brain function (D'Esposito, 2000; Detre and Floyd, 2001; Savoy, 2001). The data collected in these studies are exceedingly rich and hold potential for understanding the complex neurobiological mechanisms underlying brain systems such as memory (Cabeza *et al.*, 2002), language (Mechelli *et al.*, 2003a), and motor function (Kawato *et al.*, 2003) among many cognitive domains. Most contemporary fMRI studies routinely include multiple functional scan runs (i.e., on the order of five to ten minutes per run), the collection of high-resolution anatomical images of the brain, involve multiple levels of experimental manipulation, and increasingly involve several experimental groups. Stimulus input and behavioral output files also contribute to the body of study meta-data, 4-dimension image volumes, and images of brain structure. However, despite their being summarized and interpreted in the peer-reviewed literature, the raw image data and accompanying meta-data from these studies are often not accessible to other researchers in the neuroscientific community. This makes it difficult for others to rigorously inspect results, verify published claims, or to conduct novel analyses of these rich datasets. In the field of functional neuroimaging, despite being available for over a decade, databases of systems-level functional brain imaging studies of cognition have begun to receive increasing attention (Van Horn and Gazzaniga, 2002).

The field of neuroimaging has for some time recognized this need for archiving functional neuroimaging results (Fox and Lancaster, 1994; Fox *et al.*, 1994b). However, databasing and data sharing remain relatively unfamiliar ideas to many

in the field of functional neuroimaging. Coupled with the fact that the size of fMRI studies is growing rapidly, few tools exist that assist investigators in managing the data they collect and that facilitate efficient data sharing. The needs of the community are particularly unique and extant software models are not broadly applicable for dealing with many of the issues a large archive like this requires as well as helping users manage data at their own sites. Particular problems include: Users of databases that store experimental results are often faced with entering data retrospectively; certain models for sophisticated high-throughput data analysis (Roland *et al.*, 2001) put certain constraints on the type and form of data that can be stored; the management of study data is left up to the investigator until completion (e.g., publication) of the study; and finally, the tools available from existing databases are largely bibliographic and not entirely useful until all information about a study has been collected. In other words, software tools that help an investigator gather, manage, and share fMRI data that preserves that data's utility to the investigator beyond its being shared do not widely exist.

Simply managing these large amounts of data, however, is an increasing challenge for many in the field. Future investigations of *in vivo* brain function using fMRI can be expected to continue gaining in sophistication, as the questions being asked about the brain, and the methods themselves, become more elaborate. With possibly hundreds or thousands of data files, each recorded under conditions that vary over time, the amount of data may soon overwhelm the abilities of investigators to keep track of this information from even a single fMRI investigation. Being able to manage and organize these large and varied collections of data is a prelude to subsequent data sharing and databasing.

In this chapter, we review the form, size, and means of processing of functional neuroimaging data but highlight the potential of these data to be used beyond the initial interpretation by the original investigators; how these data may be used to develop and test new ideas about human brain function as well as for developing useful computational algorithms for extracting new results; we discuss the emerging role of high-performance and Grid computing; the use of clever visualization techniques, and the availability of online databases of neuroscientific data. Finally, we discuss some of the challenges that must be overcome for these and other resources to fully leverage the wealth of neuroscientific information into new discovery.

II. Examining Cognitive Function with fMRI

The signal of interest using positron emission tomography (PET) is based on the fact that changes in the cellular activity of the brain of normal, awake humans and laboratory animals are accompanied almost invariably by changes in local

blood flow (Raichle, 1975, 2001b; Sokoloff, 1981). Early PET studies of the brain's response to cognitive tasks provided a level of precision in the measurement of blood flow that opened up the modern era of functional human brain mapping (Raichle, 2003). Functional MRI, on the other hand, distinguishes itself from PET by capitalizing on endogenous magnetic properties of deoxygenated to oxygenated hemoglobin in order to track regional cerebral blood flow (Hoppel *et al.*, 1993; Rosen *et al.*, 1993). Using fMRI to visualize brain function *in vivo*, neuroscientists have demonstrated that the mental operations carried out by the human brain can be empirically and repeatedly measured (Bandettini and Wong, 1997; D'Esposito, 2000) and, since the early 1990s, fMRI has taken the place of PET as the most widely used method for brain mapping and studying the neural basis of human cognition. Though now enjoying widespread practice throughout the world, an incomplete understanding of the physiological basis of the fMRI signal has remained to confidently interpret the data with respect to neuronal activity. The biological origins for these signals is an area of much interest for the application of tools for cognitive neuroscience research and modeling (Raichle, 2001a; Woo and Hathout, 2001). Understanding the origins of the BOLD signal is useful for informing models of the hemodynamic response function (Buxton and Frank, 1997; Buxton *et al.*, 1998) or to guide characterization of the neurophysiological processes that occur in advance of BOLD signal change as a result of many higher-order cognitive models (Friston, 2002; Friston and Price, 2001; Price and Friston, 2002).

New insights into higher cognitive functions, such as episodic and working memory (Cabeza *et al.*, 2002; Carpenter *et al.*, 2000), linguistic processes (Binder *et al.*, 1997; Buchel *et al.*, 1998; Crosson *et al.*, 1999), and object visual processing (Beauchamp *et al.*, 2002) have been described. Face perception is one particular cognitive operation to be extensively examined using fMRI (Haxby *et al.*, 2000) and appears to be governed principally in the ventral portion of the temporal lobe—the "fusiform face area" (Kanwisher *et al.*, 1997). Brain areas bordering this region may be sensitive to the spatial properties of pictures of other objects, such as chairs and houses (Ishai *et al.*, 2000) with spatially distributed but overlapping portions (Haxby *et al.*, 2001). Additional research has indicated that this region may, in fact, be specialized for visual recognition expertise which includes processing for faces (Gauthier and Nelson, 2001; Gauthier *et al.*, 1999). Work has also investigated the social context of face perception, in particular with respect to the perception of threat (Adolphs, 2003; Haxby *et al.*, 2002; Richeson *et al.*, 2003), the familiarity of faces (Leveroni *et al.*, 2000), and the processing of faces in diseases such as autism (Adolphs *et al.*, 2001). Visuospatial attention has also been explored using fMRI (Binkofski *et al.*, 2002; Culham *et al.*, 2001; Hamalainen *et al.*, 2002; Kanwisher and Wojciulik, 2000). fMRI studies have pointed toward a network of cortical visuospatial and oculomotor control areas, specifically the lateral occipital cortex, precentral sulcus, and intraparietal sulcus, as being active

in covert shifts of spatial attention (Beauchamp *et al.*, 2001). In parietal and frontal cortical areas, BOLD activation increased with attentional load, suggesting that these areas are directly involved in attentional processes, though this was not evident in the fusiform gyrus (Culham *et al.*, 2001), indicating possibly separate but complimentary systems underlying attention to stimuli such as human faces.

A number of factors have been implicated in the origins of the BOLD response including energetics, oxygen consumption, as well as parameters such as blood volume and flow (Buxton *et al.*, 1998). The question of whether the BOLD response is the result of neuronal output or if it is due to the internal communication among localized populations of cells has also been recently addressed. Logothetis and coworkers (2001) conducted the first simultaneous intra-cortical recordings of neural signals and hemodynamic responses. Varying the temporal characteristics of the stimulus, they observed a moderate to strong association between the neural activity measured with microelectrodes and the pooled BOLD signal from around a small area near the microelectrode tips. However, the BOLD signal showed significantly higher variability than the neural activity, indicating that human fMRI coupled with traditional statistical methods underestimates the reliability of the neuronal activity. To further characterize the relative contribution of several types of neuronal signals to the hemodynamic response, they compared local field potentials (LFPs), single- and multi-unit activity (MUA) with high spatiotemporal fMRI responses recorded simultaneously in primate visual cortex. Selecting recording sites having transient responses, only the LFP signal showed significant correlation with the hemodynamic response and were superior to MUA at predicting the fMRI response. Thus, BOLD signal is a putative measure of the input and processing of neuronal information within brain foci, not the output signal transmitted to other brain areas.

Epoch or "block" experimental designs have been the work horse of fMRI experimentation and are those in which stimuli are presented for some period of seconds (several TRs or brain volume sampling intervals) and alternated randomly or pseudo-randomly over the course of the data acquisition period. They are the easiest to conduct and tend to provide robust activation in most tasks but may limit the number of stimulus types that can be presented. Conversely, event-related experimental designs are characterized by having a baseline time course that is punctuated with stimulus events. Event-related methods, conversely, have permitted a broad array of task designs to be explored with brain imaging techniques (Buckner, 1998; Buckner *et al.*, 1996; Rosen *et al.*, 1998). Individual trial events can be presented rapidly, in randomly or intermixed order, and the hemodynamic responses associated with each trial event type reliably estimated (Dale and Buckner, 1997). The basis of event-related studies is that the hemodynamic response tracks neuronal activity on the temporal scale of seconds and, in many situations, summates over trials in a manner well predicted by a linear model that is sufficient even for very briefly spaced stimuli (e.g., ∼2 seconds). With this increased

interest in event-related paradigms in fMRI, there has been considerable effort in identifying the optimal stimulus timing, especially when the inter-stimulus interval is varied during the imaging acquisition run (Birn *et al.*, 2002; Dreher *et al.*, 2002).

Experimental designs for event-related functional magnetic resonance imaging can be characterized by both their detection power, a measure of the ability to detect activation, and their estimation efficiency, a measure of the ability to estimate the shape of the hemodynamic response. Computer simulation studies have indicated that estimation of the hemodynamic response function is optimized when stimuli are frequently alternated between task and control states, having shorter interstimulus intervals and stimulus durations, while the overall detection ability of activated areas is optimized when using blocked designs (Birn *et al.*, 2002; Mohamed *et al.*, 2000). This suggests that event-related designs may provide more accurate estimates of the HRF than epoch-related designs, with the maximal response to events occurring sooner and returning to baseline later than in a stimulus epoch (Mechelli *et al.*, 2003b). The choice of data processing operations, however, can affect statistical inference in all designs and means for optimizing data processing pipelines is an area of active research (LaConte *et al.*, 2003; Lukic *et al.*, 2002; Strother *et al.*, 2002).

Functional neuroimaging using MRI promises to continue growing as the principal method for examining *in vivo* brain function. Though individual studies using fMRI promise to reveal much about such basic brain processes, there also exists great potential for contrasting, comparing, and combining these studies to explore fundamental properties of cognitive function as well as the properties of the BOLD signal itself. The large amount of information collected in an individual study, however, and how this could be mined by others to produce novel research is an under-appreciated aspect of this work that is worth addressing further. There is often more information contained in a neuroimaging study that can be adequately described in a single neuroimaging article. The expertise required to extract this information, however, is often not necessarily possessed by the original study authors. Finally, new imaging facilities are very costly and to install MRI scanners in psychology departments across the country may not be cost effective in contrast to providing an open archive of such data where researchers can readily obtain the original fMRI time series and subject them to new analyses.

III. Large-Scale Archiving of fMRI Study Data

The advent and development of fMRI has resulted in a quantum leap in the ability to visualize the brain's capabilities. However, this has also vastly increased the amount of information that brain researchers must manipulate, manage, and store. fMRI study data sets are large, often exceeding several gigabytes (GB) in

size. As advances are made in MRI scanner technology to permit the more rapid acquisition of data, functional imaging experiments will consist of more data per unit time over the same scan duration. As cognitive neuroscientists ask ever more sophisticated questions about fundamental brain processes, they will undoubtedly collect data on a greater number of subjects and more fMRI time courses per subject. Indeed, archives equivalent in size to that of several petabytes are not out of the question and will likely be the norm within the next decade. A number of individual fMRI data sets already rival the full size of many extant large genetic (Ackerman, 1999; Ackerman and Banvard, 2000) and protein (Chen and Xu, 2003; Legato *et al.*, 2003; Noguchi and Akiyama, 2003) science data archives (see Table II, below, for comparison). For instance, the complete study data from Buckner *et al.* (2000) (fMRIDC Accession#: 2-2000-1118W) represents a study in excess of 20GB. It can be expected that as technological advances are made in MRI scanner technology which improve the spatiotemporal resolution of the data obtained, the amount of brain image data collected in published articles will routinely rival the size of the human genomic database. A challenge therefore exists in devising efficient means for comparing and contrasting these data on a large-scale but within a reasonable time frame.

As fMRI use in cognitive, clinical, and social neuroscience grows and becomes more widespread, individual researchers must be prepared for the large disk storage requirements that are needed to contain the data and their analyses. A greater number of subjects, for instance, improves the inferential power of the statistical tests performed and helps researchers to be confident in the effects they observe (Van Horn *et al.*, 1998). However, increases in sample size readily require increased costs associated with each fMRI study, in terms of scanner time, subject reimbursement, among other expenses. Publicly accessible archives of these data (for example see Table I) can help spread the costs of this research over the community, whereby the researcher may perform re-analyses on existing data at a greatly reduced cost compared to collecting the data themselves.

With these issues in mind, the fMRI Data Center (fMRIDC; *http://www.fmridc.org*) was established as a public archive for fMRI study data and the associated experimental meta-data. The fMRIDC began receiving data from researchers in 2000 and began making datasets publicly available in 2001. At present, the archive contains over 100 complete data sets which researchers may request online and have shipped to them free of charge. Authors of fMRI studies have been asked to provide the details of their experiments across several levels: description of the subjects taking part in the experiment (e.g., their age, handedness, clinical diagnosis, etc.); the description of the MRI scanner (e.g., manufacturer, model, software revision, field strength, etc.) as well as the scanning session protocols used during the study (e.g., number of slices acquired, echo time (TE), relaxation time (TR), etc.); and, finally, the details of the experimental design (e.g., stimulus time course information, number of experimental

TABLE I
BRAIN DATABASE RESOURCES AVAILABLE ONLINE

Database name	Principle modality	Data sets provided	Public access?	Species	Country and funding source[a]	Web site URL
Allen Brain Atlas	Anatomical sections	Photomicrographs of gene expression	Limited	Mouse	US; Private	http://www.brainatlas.org
BIRN	MRI	Images (MRI and cell photomicrograph)	Limited; greater access to participating BIRN centers	Human, Mouse	US; NCRR	http://nbirn.net/
BRAID	MRI, fMRI	Image volume data	Limited	Human	US; N/A	http://www.rad.upenn.edu/sbla/braid/publications/all.shtml
Brain Gene Expression Map (BGEM)	Anatomical sections	Photomicrographs of gene expression	Open	Mouse	NIH/NINDS, ALSAC	http://www.stjudebgem.org/web/mainPage/mainPage.php
BrainMapDBJ	PET/fMRI	Results local maxima	Limited; greater access to participating ICBM centers	Human	US; NLM	http://www.brainmapDBJ.org
BrainWeb	MRI	Simulated MRI image volume data	Open; part of the LONI/ICBM consortium	Human	Canada; NIMH/HBP Non-US	http://www.bic.mni.mcgill.ca/brainweb/
BREDE	fMRI	Results local maxima, VRML, XML	Open	Human	Denmark; NIMH/HBP Non-US	http://hendrix.imm.dtu.dk/software/brede/

CoCoMac	Single and multi-unit recordings	Neural connectivity data	Open	Non-Human Primate	Germany; Non-US	http://www.mon-kunden.de/cocomac/
EarLab	Single/multi-unit recording	Cell recording time series	Open	Non-Human	US; NIMH/HBP	http://earlab.bu.edu/
fMRIDC	fMRI, MRI	Raw, processed, results, anatomical brain images and study meta-data	Open	Human	US; NSF, Keck, NIMH/HBP	http://www.fmridc.org
International Brain Volume Database (IBDV)	MRI	High resolution structural Image volumes	Open	Human	US; NIMH/HBP	http://www.cma.mgh.harvard.edu/ibvd/
LONI/ICBM	PET, MRI, fMRI, EEG, MEG	Image data	Limited; greater access to ICBM centers	Human	International; NIMH/HBP NCRR, Private funding	http://www.ioni.ucla.edu
Mouse Brain Library (MBL)	Anatomical sections	Photomicrographs	Open	Mouse	US; NIMH/HBP	http://www.mbl.org
Neurodatabase.org	Single/multi-unit recording	Cortical neuron electrical recordings	Open	Multiple	US; NIMH/HBP NINDS	http://neurodatabase.org
Neurogenerator	PET, fMRI	Imaging data submitted by users is organized into a database that is returned to the user	Limited	Human	Sweden; The European Commission	http://www.neurogenerator.org
SenseLab	Single/multi-unit recording	Cell recordings from multiple sources	Open	Non-Human Primate	US; NIMH/HBP	http://www.senselab.yale.edu
Surface Management System (SuMS)	MRI	Digitally-based cortical surface models	Open	Human, Non-Human Primate	US; NIMH/HBP	http://brainmap.wustl.edu/sumshome/

[a]Where evident from the database Web site

runs, etc.). Table I presents a summary of the items requested from authors which describe their experimental data. The principle intent of obtaining this degree of information about each study is that it should be complete such that another researcher could take the information and the accompanying brain image data and reconstruct the results reported in the literature by the original authors (see Van Horn *et al.*, 2001 for review).

IV. The Emergence of "Discovery Science"

The collection of biological data into large databases has led to a change in thinking about the potentially restrictive nature of strictly hypothesis-based research. Increasingly, researchers are beginning to move toward a science of discovery—examining vast and disparate collections of data and hunting for unseen patterns that might provide clues to underlying biological mechanisms. The mountains of data being collected in many fields provide input for pattern-seeking and other relevant algorithms that can provide additional insights into complex, multidimensional data (Jones and Swindells, 2002; Ma *et al.*, 2002; Schutte *et al.*, 2002). These patterns can suggest mechanisms, and the mechanisms can, in turn, suggest testable biological experiments to foster new hypothesis-driven research. Confirmed mechanisms add to the knowledge base of the biological sciences and provide the basis for further discoveries including those that will improve quality of life and provide the means for attacking disease.

Such mining of the integrated resources developed and disseminated by the NCBI, Genbank, and the Human Genome Project has led to several scientific advances. The discovery of the genes for hereditary nonpolyposis colorectal cancer (HNPCC) is one such example. HNPCC is thought to account for one-sixth of all colon cancer cases (van Stolk, 2002). Although most forms of cancer appear to be nongenetic, there are certain forms where a person has a hereditary risk attributable to a single altered gene (Calvert and Frucht, 2002). Using the tools developed through the Human Genome Project, notably Genbank, an international research team tracked the gene to a specific region of chromosome 2 (Lindblom *et al.*, 1993). Researchers then identified a second gene on chromosome 3 that was also associated with this form of cancer (Peltomaki, 1994). Together, mutations within these two genes are responsible for the majority of cases of HNPCC. Researchers have used this new knowledge to develop blood tests to screen select individuals for these gene mutations (Ramesar *et al.*, 2000; Thomas, 1994). Detecting the presence of the mutated genes for HNPCC within a family allows clinicians to target relatives most likely to benefit from treatment. By identifying an unaffected family member at risk for HNPCC, physicians may then more closely monitor them for signs of disease development. Family

members determined to be noncarriers no longer have to suffer through extensive medical examinations. Most importantly, patients demonstrating early signs of cancer and determined to carry a gene mutation may undergo prompt medical treatment. Due to the role played by informatics, when diagnosed and treated early, HNPCC is nearly 100 percent curable (Boardman, 2002).

Several such large-scale, infrastructural, and discovery-focused database research efforts that are already seeing considerable scientific payoff are underway in the biological and astrophysical sciences. The successes of these molecular biological, biomedical, and astrophysics infrastructures are well known. They have provided the means for experts in computer science, mathematics, and statistics to make significant contributions to these fields from which most of their expertise would have been excluded without the infrastructure. These successes are not necessarily unique but building upon them and extending them to a wider set of scientific research arenas is an ever-present theme (Altman, 2003; Brookes, 2001; Persson, 2000).

Before this process can begin for any particular field of science, however, an infrastructure must be laid down that will support these new approaches. The path blazed by the molecular biologists is, once again, illustrative. Likening these efforts to civil engineering projects, Eric Lander, Director of the Whitehead-MIT Center for Genome Research, has noted that programs to develop computational infrastructure represent "very important roads." (Incyte Genomics interview (2001) (http://www.incyte.com/)).

V. Data Sharing in Neuroscience

The driving force behind many biological and physical science informatics, data mining, and research initiatives has revolved around the sharing of primary research data (Becker, 2001; Ilioudis and Pangalos, 2001; Reidpath and Allotey, 2001). The National Institutes of Health (NIH) in the United States have recognized the benefits that the sharing of primary research data has for advancing science and has recently implemented policy requiring data sharing for grants in excess of US$500K/yr in direct costs (*Final NIH Data Sharing Policy Notice*: http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html).[1] Likewise, the National Science Foundation (NSF) in the US has encouraged data sharing for several years, in the social and economic sciences in particular, (*http://www.nsf.gov/sbe/ses/common/archive.htm*). The Medical Research Council (MRC;

[1]Further information on the NIH data sharing policy may be found on the NIH Data Sharing Web Page: *http://grants.nih.gov/grants/policy/data_sharing/*

http://www.mrc.ac.uk) in Great Britain also strongly encourages open scientific data sharing.

The NIH's position on coordinated scientific data sharing, particularly as it relates to neuroscience, has recently been underscored in an essay by several NIH institute directors: "Efforts driven by collaboration, coordination, and computation should yield the data, tools, and resources that neuroscientists will need in the coming decades." (Insel *et al.*, 2003). The sharing of primary research data is needed to provide a record of the scientific body of work, permit comparison of various approaches to studying brain function, and enable large-scale analyses across data sets. There are several models for the sharing of research data that form a spectrum of complexity and detail. These include models for data archive access; the simple model of an anonymous FTP site, where a data set is placed and openly available but with no guarantee that the data have not been subjected to quality control, been published in a peer-reviewed publication, or that the dataset will be maintained; peer-to-peer models, wherein individual investigators set up and personally maintain private data sharing relationships with colleagues and co-investigators of both published and unpublished data; the conforming site model, in which consortiums of several research centers agree to exchange data through conventions established and governed by one of the consortium member centers, but with no guarantee that nonconsortium members may have access to the data archive or will have access to consortium-derived software tools needed for interacting with the data; and the centralized repository model, in which complete datasets are contributed, curated, and maintained in a central site by dedicated personnel and made openly available to the entire research community.

Data sharing models also focus on the type and amount of data that should be shared. Recent commentary has suggested that the value of shared neuroimaging data is greatest only after processing has been applied and interpretation provided by study authors (Fox and Lancaster, 2002). However, the information content of the image voxel time course data remains the same or is reduced by every step of processing (Van Horn and Gazzaniga, 2004). Therefore, it is unclear as to the amount of added value when archiving only statistical local maxima tables obtained after the data are heavily processed.

The model for sharing peer-reviewed study data in which the potential benefits of a data set are likely to be greatest is when the data are curated in a centralized location by a dedicated staff, complete study data have been indexed, and are freely available to the entire community (Van Horn and Gazzaniga, 2004). Accompanied by detailed, ontologically-structured, study meta-data, and a comprehensive description of data processing methods, experiment image data may be examined by other researchers at various points in the processing chain (raw, processed, or results) depending upon the needs and interests of colleagues or independent researchers. Through centralized curation and open distribution, efforts to subject functional data to re-analysis or perform mega-analyses across

data sets may be maximally successful and thereby promote unique scientific discovery and advance education. This model of data sharing helps promote the cycle of science by adding an extra component to the publication process that may enhance new research and education, foster new avenues for research, and contribute back into the collective body of knowledge.

VI. The Role of Computation in Neuroscience

The NIH Roadmap (*http://nihroadmap.nih.gov/*) stresses the importance of computational biology, bioinformatics, and the establishment of digital science libraries. A recent NSF Blue Ribbon Advisory Report emphasizes that computers, computer science, and technology are at the heart of the future of a range of research fields that have "profound broader implications for education, commerce, and social good." (*http://www.communitytechnology.org/nsf_ci_report/*, (Blue Ribbon Advisory Panel On Cyberinfrastructure, 2003)). Though perhaps overdue in fully recognizing the potential of computers and the internet, the field of neuroscience is now growing in its dependence on high-end computational infrastructure. Neuroscience, in particular cognitive neuroscience, has emerged over the past decade as a cross-cutting aggregate of these key areas with an emphasis on the human brain information mining, modeling, and visualization (Adolphs, 2003; Casey, 2002; Corchs and Deco, 2002; Toga, 2002a). To broaden participation in understanding brain function derived from technologies such as brain imaging, greater reliance upon computational infrastructure to facilitate research collaboration is required. The sharing of large data sets via the internet; being able to collect, archive, and index these data; and subject these data to high-throughput analysis is not an option, but a mandatory next step in the advancement of understanding of normal brain function. Via this route, large, culturally-, and gender-valid norms must be established, for example, forming a benchmark against which to provide diagnosis in brain illness and disease.

Several novel concepts have been borne out of this interest in large-scale, scientific collaborative infrastructure. Most notably is the concept of Grid computing (Butler, 2003), the basis of which has existed for several years in the form of distributed computing, and the emerging need for high-speed Internet connectivity. The middleware for the development of global interconnected computer systems has made great progress in the last few years (Avery, 2002). The Grid software enables users, tools, and computer hardware to interact and share resources over high speed connections in an Internet- and Internet standards-compliant fashion (Foster, 2003). Users of Grid-enabled systems will be able to write applications to these published interfaces and will expect to be able to run on large-scale heterogeneous systems (Fig. 1). The Web, and in particular

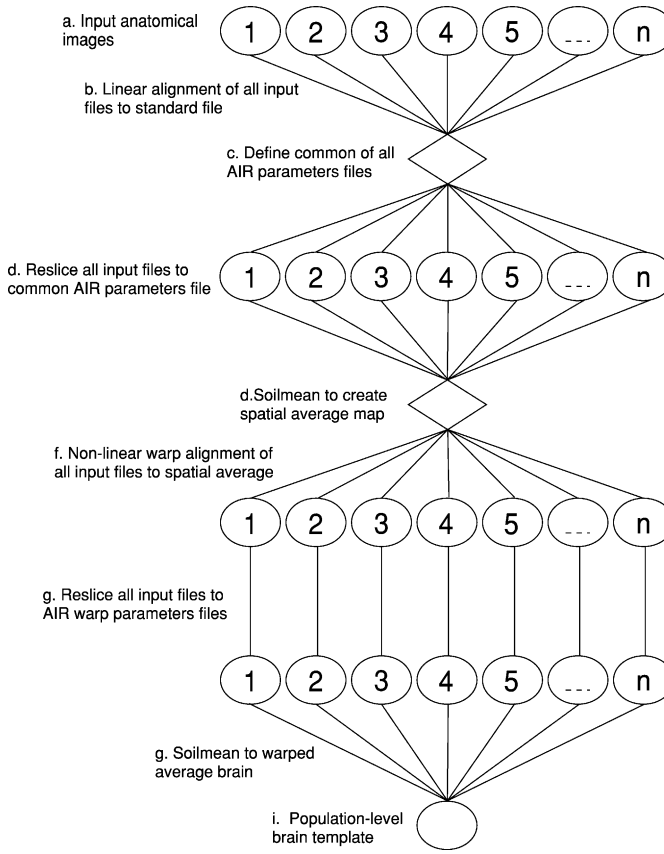

FIG. 1. Neuroimaging data processing pipelines are ideally suited to Grid-based distributed computing.

Web – services, can provide a model for this large-scale computational system. Various fields have been earlier adopters of Grid computing, notably the High Energy Physics community. Several worldwide and multi-institutional Grid projects are underway to enable simulations on large distributed datasets. The TeraGrid (*http://www.teragrid.org*) is a NSF-funded project to promote distributed scientific computing using Grid infrastructure to connect the nation's largest supercomputer centers. Using the Grid as a backbone, The Globus Toolkit (http://www.globus.org) has become the *de facto* standard for Global Grid communications. The Globus Alliance leads the development of the toolkit. The Global Grid Forum (GGF) is currently leading the standards efforts which Globus implements. Shared computational infrastructure, tools, and tool development, as well as collaborative research on archival data leading to new, testable

hypotheses is becoming easier each day. These technological advances in large-scale computing on shared infrastructure have important implications for neuroscientific research dealing with massive amounts of data as is the case in functional neuroimaging.

VII. Brain Data Repositories as a Shared Resource for Neuroscience

When the fMRI Data Center effort was initiated, one of the central intents was to build an Internet-accessible platform through which researchers based at other institutions might access the growing collection of fMRI study data to evaluate methodologies for data processing (LaConte *et al.*, 2003; Lukic *et al.*, 2002); guide the design open-source software tools for data management (Van Horn *et al.*, 2002); construct means to summarize these large data sets to facilitate rapid search, visualization, and discovery; all with a view toward driving new hypothesis-based fMRI research. The fMRIDC team, in particular, has worked to construct a shared community access cluster Grid system for the analysis of functional datasets from the fMRIDC archive. This system has seen increased utilization as an analysis platform to mitigate the current CD delivery of datasets. Accounts on the system are available to any member of the fMRI community who wishes to perform a large analysis of data from the fMRIDC (see *http://www.fmridc.org/grid*).

The software infrastructure provided by Grid services are needed to perform meta-analyses on the multi-gigabyte fMRI datasets housed in the fMRIDC archive. The fMRIDC has implemented a number of components from the Globus Toolkit 3.0, including the Grid-FTP service. Current plans exist for the full toolkit to be installed to enable multi-institutional scheduling of resources. The Globus software will abstract the fMRIDC's own Grid scheduling system and the systems of our collaborators as well as those of other systems. With advanced computational resources dedicated to neuroimaging and that are deployed using Globus, new methods can be developed through testing on what is fast becoming the world's largest data warehouse of functional neuroimaging data.

VIII. fMRI Data Archiving, Mining, and Visualization

Like much of neuroscience, literature-driven, hypothesis-based approaches have been the underlying approach to most cognitive neuroscientific investigations over the past 100 years. However, streamlined, computationally efficient approaches to examining large amounts of data are emerging as advantageous where volumes of information from diverse sources impede a straightforward test

of experimental hypotheses (Baumgartner *et al.*, 2000; Friman *et al.*, 2002; Jarmasz and Somorjai, 2002). Approaches used in the biological sciences include nonparametric clustering methods (Cordes *et al.*, 2002; Goutte *et al.*, 2001; Salli *et al.*, 2001), pattern searching (Cummings *et al.*, 2002; Jones and Swindells, 2002), as well as novel visualization techniques (Baumgartner and Somorjai, 2001; Teo *et al.*, 1997). In brain imaging, similar methods might be used on a database as a precursor to more thorough parametric hypothesis testing using more sophisticated modeling methods (Cox, 1996; Cox and Hyde, 1997; Friston *et al.*, 2002; Lohmann *et al.*, 2001) on subsets of the overall study space. Exploratory data analysis approaches (Tukey, 1977; Velleman, 1981) often assume little about the underlying data, which allows the data to more freely inform the investigator about itself using more elementary statistical approaches than might be true if the data were assessed using highly parameterized modeling procedures. These methods have been successfully applied to fMRI data with promising results (Baumgartner and Somorjai, 2001; Baumgartner *et al.*, 2000; Friman *et al.*, 2002; McKeown and Sejnowski, 1998). They can often be used to highlight relationships among the overall collection of data in ways not possible when viewing only a single data set. These relationships may not emerge until viewed in the light of a large number of other studies to which they may be compared. Thus, neuroscience, in particular studies using fMRI, has a great potential to become a discovery-based science, where exploratory analyses can lead to new ideas worth pursuing with hypotheses-generated experimentation (Van Horn and Gazzaniga, 2002). But to put such ideas into practice, however, requires an interdisciplinary approach, bringing together experts in cognitive, computer, and mathematical science to work jointly in solving the challenges inherent in large-data science.

Initiatives to archive neuroscience data form a unique collaboration between cognitive neuroscientists, mathematicians, and computer scientists to explore the challenges inherent in (1) data warehousing—*en mass* data storage; (2) data mining—how to apply efficient mathematical and computer algorithms to sift through large amounts of data to extract unique and interesting features from fMRI study data; and (3) data distribution—effective means to permit others to interact with the data archive. For example, the efficient retrieval of useful information from these large datasets poses many interesting problems for which computer scientists play an important role in providing answers. Methods for applying sophisticated search queries across multiple levels of neuroimaging study data (Table II) that could be investigated include: (1) searches for key text phrases across the published research article itself (i.e., the PDF version of the published study) and subsequent document clustering by assessing the usage of similar words at similar rates; (2) queries across the study "meta-data" composed of scanner protocol, experimental paradigm, subject demographic, and other information provided by the study authors; and (3) the 4D fMRI image time course data itself by, for instance, performing image timecourse-based clustering

TABLE II
BASIC fMRI STUDY INFORMATION COLLECTED FOR THE fMRIDC ARCHIVE

MR Scanner Protocol Information

- Scanner Protocol ID
- Scanner Head Coil Type
- Pulse Sequence Type
- Flip Angle (degrees)
- TE (in milliseconds)
- TR (in milliseconds)
- Number of time-points
- Number of acquisitions
- Number of dummy scans
- Number of slices
- Slice thickness (in millimeters)
- Slice skip (in millimeters)
- Interleaved or sequential slice acquisition
- Field of View (FOV)
- Receiver bandwidth (MHz)
- Original image acquisition matrix size
- Reconstructed image acquisition matrix size
- Full or partial K-space
- Image Acquisition Orientation
- Ramp sampling
- Echo train length
- Echo shift in asymmetric spin-echo
- Type of reference scan for reconstruction

Subject Information

- Subject ID
- Experimental group code
- Gender
- Age
- Health Status
- Assessments (e.g., handedness, etc.)
- Medication status
- Other (e.g., diagnostic, etc.)

Scan Session Information

- Scan Session ID
- Scanner Manufacturer
- Scanner Model
- Scanner software revision
- Magnet field strength
- Scanner Gradient Slew rate
- Date of scan session
- Duration of scan session
- Other

(*Continued*)

TABLE II (*Continued*)

Experimental Protocol
• Experimental Protocol ID
• Number of groups
• Number of subjects per group
• Number of functional runs
• Epoch-related conditions
• Event-related conditions
• Experimental methods
• Stimulus regressor files
• Other (e.g., additional condition descriptions, associated data files, etc.)

accompanied by *a posteriori* probabilistic classifier algorithm to measure classification reliability. The latter of these approaches has necessitated the application of summarizing signal processing and information theoretic methods for rapidly analyzing large fMRI data sets. The integration of these levels using leading-edge, computer-based IR algorithms permit "global" study clustering in order to "learn" what is needed to identify interesting patterns within and between these levels of data. The application of algorithms that permit fMRI data self-description, allowing the data to tell an investigator about itself rather than through the fitting of statistical models, as is common in functional neuroimaging, is a promising application of "machine learning" (Mitchell, 1997; Mitchell, 1999).

Another area of active interest in neuroimaging data representation lies in identifying unique approaches to the visualization of this massive amount of information. Traditional approaches to visualizing brain imaging study results have relied on overlay patterns of brain activation from functional scans on top of high-resolution structural images (Fig. 2). Popular methods of display include representing patterns of functional activity on flattened models of the cortical surface (Van Essen *et al.*, 2001a; Van Essen *et al.*, 2001b). Still other, novel methods for displaying and interacting with more abstract representations of large collections of information are needed that may reveal previously unseen relationships in the data. For instance, approaches centering on nontraditional and abstract methods of data exploration, such as taking the data out of "brain space" and placing it in some alternative parameter space, and examining patterns in the data that might have been invisible in the original anatomically-based space (Baumgartner and Somorjai, 2001; Cordes *et al.*, 2002). These may also include iconic representations of data endowed with synthetic physical properties that distort the relative coordinates of data in an abstracted space. Such models have been successfully employed in visualizing biochemical and metabolic pathways (Becker and Rojas, 2001; Karp and Paley, 1994; Ogata *et al.*, 2000). Coherent subspaces within this abstracted space can be identified and examined as those where something of neuropsychological interest might be

FIG. 2. The novel visualization of fMRI results can enhance or emphasize areas not fully appreciated in the original published article. This figure shows a VRML rendering, exported from the Caret software package, of the memory encoding statistical main effects from the study of Druzgal and D'Esposito (2001) viewed as interactive object using the Cortona Client for MS Windows XP, Version 4.2.R93, from Parallel Graphics, Inc. (*http://www.parallelgraphics.com*). This rendering highlights the robust activation of the fusiform gyrus, the superior portion of the anterior cingulate, inferior parietal lobule, as well as the temporal pole in the right cerebral hemisphere. Viewed as a virtual reality object would enable a student or researcher to interactively translate, rotate, or zoom in on regions of particular interest and to view the data as a whole, rather than as a collection of 2D overlay graphics.

occurring. As fMRI archives continue to grow through active data sharing, working with the shear amount of data, as well as its direct visualization, becomes increasingly difficult. These computational and visualization methods will allow large amounts of data to be processed and visualized.

IX. Neuroinformatics—The Nexus of Brain, Computational, and Computer Sciences

Given recent success stories from the domains of genomics (Escribano and Coca-Prados, 2002; Feolo *et al.*, 2000; Rafalski *et al.*, 1998) and proteomics (Berman *et al.*, 2000; Ezzell, 2002; Persson, 2000) for organizing large amounts of data, cognitive neuroscientists are likewise becoming intimately familiar with large-scale data analysis, applying high performance computing systems, and using sophisticated computer science to extract information from large archives of neurophysiological data. The evolution of the cognitive neuroscience field is fast approaching the time when it forms a confluence of brain science, high-performance computing systems, and leading edge computer science (Beltrame and Koslow, 1999; Wong and Koslow, 2001). As such, a more thorough understanding of the brain and its cognitive processes will necessitate increased computational infrastructure, novel software technology to accelerate data analysis and to mine vastly larger amounts of data, and the sharing of primary research data. Moreover, these data must be understood on a level that permits the representation of the dynamic examination of brain data and brain systems required for cognitive processes such as memory function, visual abilities, and motor skill. This effort must reach beyond the level of the examination of individual loci of brain activity to that of identifying patterns of activity across individuals that speak to the dynamics and complexity of the neural processes that are not typically reported in the scientific literature though may be worthy of additional scrutiny and study.

In response, the field of neuroscience is rapidly moving beyond its roots as a theoretical and experimental science toward becoming a highly computational science ever more dependent upon lead edge technologies in computer science, engineering, and mathematics. This is the origin of *neuroinformatics*, a unifying discipline at the nexus of information technology, computer science, and the neurosciences. It also involves the incorporation of high performance computing, visualization, and data mining techniques with the fundamentals of experimental design, image processing, and spatial and temporal statistics for neurophysiological data, in particular, for functional neuroimaging (Douglas *et al.*, 1996; Smaglik, 2000; Young and Scannell, 2000). By using computers to organize, link, analyze, and examine large, complex sets of neuroscientific data, raw data may be converted into meaningful knowledge that can be used for further experimentation

into cognitive function and the treatment of patients with neurological and neuropsychiatric disease (Beltrame and Koslow, 1999).

The Human Brain Project (HBP) funds many of the current database and neuroinformatics efforts. The HBP is a broad-based initiative which supports research and development of advanced technologies and infrastructure support through cooperative efforts among neuroscientists and information scientists (computer scientists, physicists, mathematicians, and engineers) (Brinkley and Rosse, 2002; Shepherd *et al.*, 1998). The principle aim of the NIMH-based HBP is to guide the production of new digital capabilities that provide Internet-driven information management systems in the form of interoperable databases and associated neuroscience data management tools (Shepherd *et al.*, 1998). Such software tools include graphical interfaces, querying and mining approaches, information retrieval, statistical analysis, visualization and manipulation, integrating tools for data analysis, biological modeling and simulation, and tools for electronic collaboration. The effort strongly supports open data sharing, believing that the primary data from neuroscience investigations has continued value to the field long after its initial publication (Huerta and Koslow, 1996; Koslow, 2000; Koslow, 2002). The HBP seeks to make neuroinformatics efforts funded under its auspices interoperable with other databases, tools, and centers, similar to some genomic and protein databases, and, thereby, create the capability to explore brain functional and structural interactions in even greater detail. The HBP also encourages researchers to leverage the emerging Internet capabilities for opening novel channels of communication and collaboration between geographically distinct sites.

A number of neuroscience databases exist that provide a variety of information and data pertaining to neural function. For neuroimaging, in addition to the fMRIDC, two other notable efforts exist: (1) The BrainMapDBJ database from the University of South Texas Health Sciences Center (USTHSC), and its current incarnation BrainMapDBJ (*http://www.brainmapdbj.org*),—pioneering efforts to provide access to the human brain-mapping literature and its results-based data in a manner to promote quantitative meta-analysis of related studies (Fox *et al.*, 1994a). It is comprised of a multi-level indexing scheme describing the study of experimental protocol as well as the derived Talairach-normalized local maxima from brain-mapping studies. However, raw, processed, or results image data are not provided to users of the database. Submissions to BrainMapDBJ are voluntary but peer-reviewed by the database editorial board, independently of the journal peer-review process, to review content suitability and correct coding. Access to the database is open to the public but its contents are limited only to the provided meta-data and the reported study local maxima. (2) The UCLA Laboratory of Neuroimaging (LONI) structural image database (Toga, 2002b) and the International Consortium of Brain Mapping (ICBM) Probabilistic Brain Atlas (Mazziotta *et al.*, 2001) have been constructed to provide a rigorous means for

data archiving and protection of collaborator-collected image data (*http://www.loni.ucla.edu/; http://www.loni.ucla.edu/ICBM/index.html*). Database query mechanisms ensure that no image data or identifying patient information is accessible to the public or to any others without the appropriate authorization and the expressed permission to release data from the ICBM collaborator that acquired and provided the data. The LONI database provides an integrated access and security mechanism such that perusal through archives is organized by authorized scientific groups within each ICBM laboratory. LONI maintains a large-scale computer infrastructure to maintain this archive and for use in pipelined processing of data (Rex *et al.*, 2003). Each of these efforts is a rich resource for finding information about brain function from both published and unpublished neuroimaging studies at multiple levels of detail. However, access to complete study information may be limited either due to proprietary restrictions from study investigators or through a limited scope of the data that is available.

The fMRIDC effort has helped to catalyze the field in considering the benefits of neuroscientific databasing and its potential for further advancing progress in understanding cognitive function through the open sharing of these large data sets (Van Horn, 2002; Van Horn and Gazzaniga, 2002; Van Horn *et al.*, 2001). The fMRIDC has committed to a policy of open science and provides its archive contents and software to the scientific community free of charge. The neuroscientific outcomes derived from the services offered by the fMRIDC via this novel computational resource are now beginning to bear fruit in the peer-reviewed scientific literature (Table III) and add markedly to the knowledge base of brain research.

X. Current Challenges for Neuroscience Databases

As with many neuroscience data archives, the challenge they are now presented with is how to best utilize the information contained in their database toward novel scientific outcomes that could not have existed without that large collection of data. Finding useful, rigorous, and timely answers to these and other questions will serve to demonstrate the promise of large-scale databasing and neuroinformatics methods and their utility in the study of brain function. In coming years, it can be expected that database-driven research will, indeed, help to supplement hypothesis-based experimentation, spur the formation of novel lines of research, and help to educate the next generation of neuroscientists.

It should be recognized, however, that a single online resource will not be capable of organizing and indexing all possible types of brain data. By linking information from one online resource with that contained in another, the wealth and richness of information provided by them both is increased. These linkages need

TABLE III
RECENT REANALYSES OF fMRI STUDY DATA

New authors	New journal	Purpose of new analysis	Original dataset reference (fMRIDC accession number)
Carlson *et al.*	J. Cog. Neuro., (2003)	Used canonical discriminant analysis to examine object categories	Ishai *et al.* (2000) JOCN, 12 Suppl 2, 35–51
Greicius and Menon	J. Cog. Neuro., (2004)	Used ICA to assess default-mode activity in auditory processing	Laurienti *et al.* (2002) JOCN, 14(3), 420–429
Greicius *et al.*	Proc. Nat. Acad. Sci., (2004)	Used ICA to assess alterations in default-mode activity in normal, older, and demented subjects	Buckner *et al.* (2000) JOCN, 12 Suppl 2, 24–34
Liou *et al.*	J. Cog. Neuro., (2003)	Characterize the statistical reproducibility of fMRI block design results	Ishai *et al.* (2000) JOCN, 12 Suppl 2, 35–51
Lloyd	J. Cog. Neuro., (2002)	Data assessed for patterns relevant to human consciousness	Ishai *et al.* (2000) JOCN, 12 Suppl 2, 35–51 Hazeltine, Poldrack, and Gabrieli (2000) JOCN, 12 Suppl 2, 118–129 Postle *et al.* (2000) JOCN, 12 Suppl 2, 2–24 Mechelli *et al.* (2000) JOCN, 12 Suppl 2, 145–156
Mechelli *et al.*	J. Cog. Neuro., (2003c)	To assess functional connectivity using dynamic causal modeling	Ishai *et al.* (2000) JOCN, 12 Suppl 2, 35–51
Penny *et al.*	Neuroimage, (2004)	Data used to compare dynamic causal models	Ishai *et al.* (2000) JOCN, 12 Suppl 2, 35–51

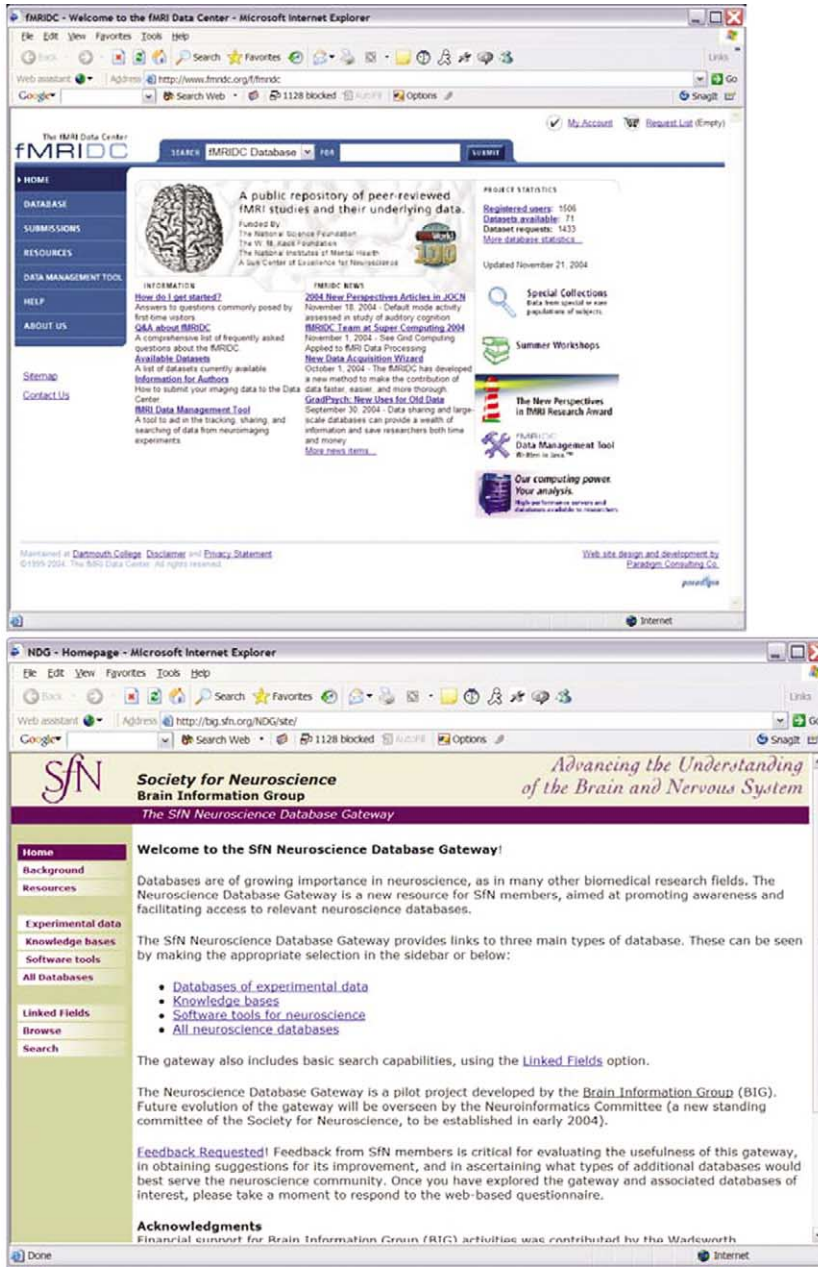

FIG. 3. (*Continued*)

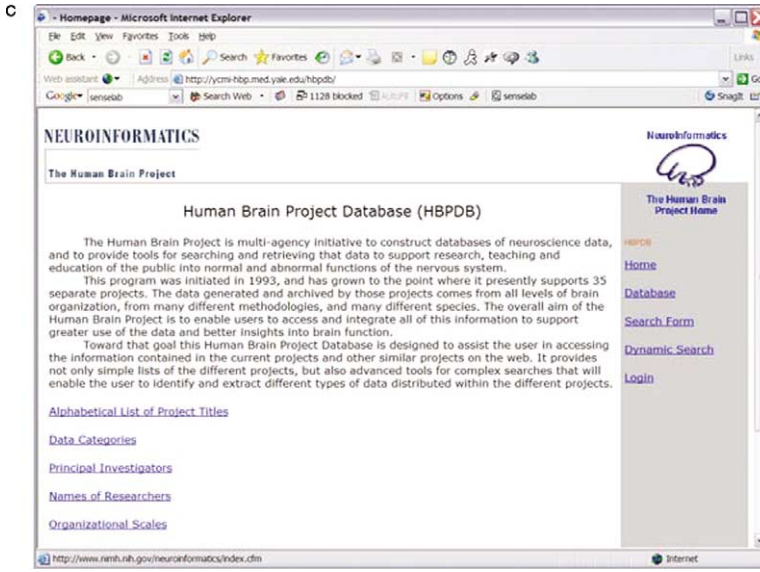

FIG. 3. (a) The fMRI Data Center web site permits researchers to browse and request complete fMRI data sets from peer reviewed, published journal articles; (b) The Society for Neuroscience provides an ever increasing portal to online neuroscience data archives (http://web.sfn.org/content/Programs/NeuroscienceDatabaseGateway/index.html), from molecular- to systems-levels, in which workers may obtain data for novel analysis and visualization; and (c) The NIMH Human Brain Project actively promotes the sharing of primary research data and lists the database efforts of investigators funded under its program (http://ycmi-hbp.med.yale.edu/hbpdb/).

not be part of a strictly federalized scheme of database participation but should span multiple independent archiving and data sharing efforts. In time, far-reaching linkages between individual resources, encouraged by the governing societies and organizations in neuroscience, will form a dynamic web of brain-related information spanning multiple temporal and spatial scales. "This notion is now being recognized by numerous funding awarding bodies. From their point of view, the sharing of primary data is now an integral part of science funding (for example, see Fig. 3)."

How the large amounts of data obtained in neuroscience experiments is best organized is another area of active interest. Ontologies have the advantage over other database frameworks that they have been developed to handle and search over qualitative information—often generally referred to as "knowledge"—as easily as the more traditional formats deal with quantitative information (Hendler, 2003). Knowledge bases, which are databases organized according to an ontology (Oliver *et al.*, 2002), rather than a strictly relational database schema, expose a middle ground between very loose and very rigid data architectures. However, they must possess the structure required for data re-use and sharing,

while maintaining the flexibility required to accommodate variations from lab to lab, researcher to researcher, and as the field concerned evolves. This approach makes sense in a context of pre-existing data management tools that need merely to be interconnected.

XI. Conclusion

Neuroscience databases are a rapidly growing resource for scientific discovery whose role in everyday neuroscience can be expected to increase in coming years. These rich archives of physiological data, brain images, genomic information, and behavioral assessments can be mined by students wishing to leverage existing knowledge into new hypotheses or used by established investigators to explore unforeseen relationships not discussed in the original published research article. Linking these resources, thereby permitting an ever denser, more enriched collection of scientific knowledge, will serve to promote and enhance brain sciences by leveraging our previous understanding toward the collection of new and exciting knowledge about brain function.

References

Ackerman, M. J. (1999). The Visible Human Project: A resource for education. *Acad. Med.* **74,** 667–670.

Ackerman, M. J., and Banvard, R. A. (2000). Imaging outcomes from the national library of medicine's visible human project. *Comput. Med. Imaging Graph.* **24,** 125–126.

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.* **4,** 165–178.

Adolphs, R., Sears, L., and Piven, J. (2001). Abnormal processing of social information from faces in autism. *J. Cogn. Neurosci.* **13,** 232–240.

Altman, R. B. (2003). The expanding scope of bioinformatics: Sequence analysis and beyond. *Heredity* **90,** 345.

Avery, P. (2002). Data Grids: A new computational infrastructure for data-intensive science. *Philos. Transact Ser. A Math. Phys. Eng. Sci.* **360,** 1191–1209.

Bandettini, P. A., and Wong, E. C. (1997). Magnetic resonance imaging of human brain function. Principles, practicalities, and possibilities. *Neurosurg. Clin. N Am.* **8,** 345–371.

Baumgartner, R., Ryner, L., Richter, W., Summers, R., Jarmasz, M., and Somorjai, R. (2000). Comparison of two exploratory data analysis methods for fMRI: Fuzzy clustering vs. principal component analysis. *Magn. Reson. Imaging* **18,** 89–94.

Baumgartner, R., and Somorjai, R. (2001). Graphical display of fMRI data: Visualizing multidimensional space. *Magn. Reson. Imaging* **19,** 283–286.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., and Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* **34,** 149–159.

Beauchamp, M. S., Petit, L., Ellmore, T. M., Ingeholm, J., and Haxby, J. V. (2001). A parametric fMRI study of overt and covert shifts of visuospatial attention. *Neuroimage* **14,** 310–321.

Becker, K. G. (2001). The sharing of cDNA microarray data. *Nat. Rev. Neurosci.* **2,** 438–440.

Becker, M. Y., and Rojas, I. (2001). A graph layout algorithm for drawing metabolic pathways. *Bioinformatics* **17,** 461–467.

Beltrame, F., and Koslow, S. H. (1999). Neuroinformatics as a megascience issue. *IEEE Trans. Inf. Technol. Biomed.* **3,** 239–240.

Berman, H. M., Westbrook, J., Feng, Z., Gilliand, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Res.* **28,** 235–242.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., and Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* **17,** 353–362.

Binkofski, F., Fink, G. R., Geyer, S., Buccino, G., Gruber, O., Shah, N. J., Taylor, J. G., Seitz, R. J., Zilles, K., and Freund, H. J. (2002). Neural activity in human primary motor cortex areas 4a and 4p is modulated differentially by attention to action. *J. Neurophysiol.* **88,** 514–519.

Birn, R. M., Cox, R. W., and Bandettini, P. A. (2002). Detection versus estimation in event-related fMRI: Choosing the optimal stimulus timing. *Neuroimage* **15,** 252–264.

Blue Ribbon Advisory Panel On Cyberinfrastructure (2003). Revolutionizing Science and Engineering Through Cyberinfrastructure. The National Science Foundation, Washington DC.

Boardman, L. A. (2002). Heritable colorectal cancer syndromes: Recognition and preventive management. *Gastroenterol Clin. North Am.* **31,** 1107–1131.

Brinkley, J. F., and Rosse, C. (2002). Imaging and the Human Brain Project: A review. *Methods Inf. Med.* **41,** 245–260.

Brookes, A. J. (2001). Rethinking genetic strategies to study complex diseases. *Trends Mol. Med.* **7,** 512–516.

Buchel, C., Price, C., and Friston, K. (1998). A multimodal language region in the ventral visual pathway. *Nature* **394,** 274–277.

Buckner, R. L. (1998). Event-related fMRI and the hemodynamic response. *Hum. Brain. Mapp.* **6,** 373–377.

Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., and Rosen, B. R. (1996). Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* **93,** 14878–14883.

Buckner, R. L., Snyder, A. Z., Sanders, A. L., Raichle, M. E., and Morris, J. C. (2000). Functional brain imaging of young, nondemented, and demented older adults. *J. Cogn. Neurosci.* **12,** 24–34.

Butler, D. (2003). The Grid: Tomorrow's computing today. *Nature* **422,** 799–800.

Buxton, R. B., and Frank, L. R. (1997). A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J. Cereb. Blood Flow Metab.* **17,** 64–72.

Buxton, R. B., Wong, E. C., and Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magn. Reson. Med.* **39,** 855–864.

Cabeza, R., Dolcos, F., Graham, R., and Nyberg, L. (2002). Similarities and differences in the neural correlates of episodic memory retrieval and working memory. *Neuroimage* **16,** 317–330.

Calvert, P. M., and Frucht, H. (2002). The genetics of colorectal cancer. *Ann Intern Med* **137,** 603–612.

Carlson, T. A., Schrater, P., and He, S. (2003). Patterns of activity in the categorical representations of objects. *J. Cogn. Neurosci.* **15,** 704–717.

Carpenter, P. A., Just, M. A., and Reichle, E. D. (2000). Working memory and executive function: Evidence from neuroimaging. *Curr. Opin. Neurobiol.* **10,** 195–199.

Casey, B. J. (2002). Neuroscience. Windows into the human brain. *Science* **296,** 1408–1409.

Chen, Y., and Xu, D. (2003). Computational analyses of high-throughput protein–protein interaction data. *Curr. Protein Pept. Sci.* **4,** 159–181.

Collins, F. S., and Mansoura, M. K. (2001). The human genome project. *Cancer* **91,** 221–225.

Corchs, S., and Deco, G. (2002). Large-scale neural model for visual attention: Integration of experimental single-cell and fMRI data. *Cereb. Cortex.* **12,** 339–348.

Cordes, D., Haughton, V., Carew, J. D., Arfanakis, K., and Maravilla, K. (2002). Hierarchical clustering to measure connectivity in fMRI resting-state data. *Magn. Reson. Imaging.* **20,** 305–317.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29,** 162–173.

Cox, R. W., and Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR Biomed.* **10,** 171–178.

Crosson, B., Rao, S. M., Woodley, S. J., Rosen, A. C., Bobholz, J. A., Mayer, A., Cunningham, J. M., Hammeke, T. A., Fuller, S. A., Binder, J. R., Cox, R. W., and Stein, E. A. (1999). Mapping of semantic, phonological, and orthographic verbal working memory in normal adults with functional magnetic resonance imaging. *Neuropsychology* **13,** 171–187.

Culham, J. C., Cavanagh, P., and Kanwisher, N. G. (2001). Attention response functions: Characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron* **32,** 737–745.

Cummings, L., Riley, L., Black, L., Souvorov, A., Resenchuk, S., Dondoshansky, I., and Tatusova, T. (2002). Genomic BLAST: Custom-defined virtual databases for complete and unfinished genomes. *FEMS Microbiol. Lett.* **216,** 133–138.

Dale, A. M., and Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping* **5,** 329–340.

D'Esposito, M. (2000). Functional neuroimaging of cognition. *Semin Neurol* **20,** 487–498.

Detre, J. A., and Floyd, T. F. (2001). Functional MRI and its applications to the clinical neurosciences. *Neuroscientist* **7,** 64–79.

Douglas, R., Mahowald, M., and Martin, K. (1996). Neuroinformatics as explanatory neuroscience. *Neuroimage* **4,** S25–S28.

Dreher, J. C., Koechlin, E., Ali, S. O., and Grafman, J. (2002). The roles of timing and task order during task switching. *Neuroimage* **17,** 95–109.

Escribano, J., and Coca-Prados, M. (2002). Bioinformatics and reanalysis of subtracted expressed sequence tags from the human ciliary body: Identification of novel biological functions. *Mol. Vis.* **8,** 315–332.

Ezzell, C. (2002). Proteins Rule. *Scientific American.* **286,** 40–47.

Feolo, M., Helmberg, W., Sherry, S., and Maglott, D. R. (2000). NCBI genetic resources supporting immunogenetic research. *Rev. Immunogenet.* **2,** 461–467.

Foster, I. (2003). The grid: Computing without bounds. *Sci. Am.* **288,** 78–85.

Fox, P., and Lancaster, J. (2002). Mapping context and content: The BrainMap model. *Nature Reviews Neuroscience* **3,** 319–321.

Fox, P. T., and Lancaster, J. L. (1994). Neuroscience on the net. *Science* **266,** 994–996.

Fox, P. T., Mikiten, S., Davis, G., and Lancaster, J. (1994a). BrainMap: A database of human function brain mapping. *In* "Functional Neuroimaging Technical Foundations" (R. W. Thatcher, M. Hallett, T. Zeffiro, E. R. John, and M. Heurta, Eds.), pp. 95–105. Academic Press, San Diego.

Fox, P. T., Mikiten, S., Davis, G., and Lancaster, J. L. (1994b). Brain-Map: A database of human functional brain mapping, pp. 95–105. Academic Press, San Diego.

Friman, O., Borga, M., Lundberg, P., and Knutsson, H. (2002). Exploratory fMRI analysis by autocorrelation maximization. *Neuroimage* **16,** 454–464.

Friston, K. (2002). Beyond phrenology: What can neuroimaging tell us about distributed circuitry? *Annu. Rev. Neurosci.* **25,** 221–250.

Friston, K. J., Glaser, D. E., Henson, R. N., Kiebel, S., Phillips, C., and Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: Applications. *Neuroimage* **16,** 484–512.

Friston, K. J., and Price, C. J. (2001). Generative models, brain function and neuroimaging. *Scand. J. Psychol.* **42,** 167–177.

Gauthier, I., and Nelson, C. A. (2001). The development of face expertise. *Curr. Opin. Neurobiol.* **11,** 219–224.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience.* **2,** 568–573.

Goutte, C., Hansen, L. K., Liptrot, M. G., and Rostrup, E. (2001). Feature-space clustering for fMRI meta-analysis. *Hum. Brain Mapp.* **13,** 165–183.

Greicius, M. D., and Menon, V. (2004). Default-mode activity during a passive sensory task: Uncoupled from deactivation but impacting activation. *J. Cogn. Neurosci.* **16,** 1484–1492.

Greicius, M. D., Srivastava, G., Reiss, A. L., and Menon, V. (2004). Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI. *Proc. Natl. Acad. Sci. USA* **101,** 4637–4642.

Hamalainen, H., Hiltunen, J., and Titievskaja, I. (2002). Activation of somatosensory cortical areas varies with attentional state: An fMRI study. *Behav. Brain Res.* **135,** 159.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293,** 2425–2430.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* **4,** 223–233.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biol. Psychiatry* **51,** 59–67.

Hazeltine, E., Poldrack, R., and Gabrieli, J. D. (2000). Neural activation during response competition. *J. Cogn. Neurosci.* **12,** 118–129.

Hendler, J. (2003). COMMUNICATION: Enhanced: Science and the Semantic Web. *Science* **299,** 520–521.

Hoppel, B. E., Weisskoff, R. M., Thulborn, K. R., Moore, J. B., Kwong, K. K., and Rosen, B. R. (1993). Measurement of regional blood oxygenation and cerebral hemodynamics. *Magn. Reson. Med.* **30,** 715–723.

Huerta, M. F., and Koslow, S. H. (1996). Neuroinformatics: Opportunities across disciplinary and national borders. *Neuroimage* **4,** S4–S6.

Ilioudis, C., and Pangalos, G. (2001). A framework for an institutional high level security policy for the processing of medical data and their transmission through the Internet. *J. Med. Internet Res.* **3,** E14.

Insel, T. R., Volkow, N. D., Li, T-K., Battey, J. F., and Landis, S. C. (2003). Neuroscience Networks: Data-sharing in an Information Age. *Public Library of Science: Biology* **1,** 9–11.

Ishai, A., Ungerleider, L. G., Martin, A., and Haxby, J. V. (2000). The representation of objects in the human occipital and temporal cortex. *J. Cogn. Neurosci.* **12,** 35–51.

Jarmasz, M., and Somorjai, R. L. (2002). Exploring regions of interest with cluster analysis (EROICA) using a spectral peak statistic for selecting and testing the significance of fMRI activation time-series. *Artif. Intell. Med.* **25,** 45–67.

Jones, D. T., and Swindells, M. B. (2002). Getting the most from PSI-BLAST. *Trends Biochem. Sci.* **27,** 161–164.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17,** 4302–4311.

Kanwisher, N., and Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nat. Rev. Neurosci.* **1,** 91–100.

Karp, P. D., Paley, S., and Romero, P. (2002). The Pathway Tools software. *Bioinformatics* **18,** S225–S232.

Karp, P. D., and Paley, S. M. (1994). Representations of metabolic knowledge: Pathways. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2,** 203–211.

Kawato, M., Kuroda, T., Imamizu, H., Nakano, E., Miyauchi, S., and Yoshioka, T. (2003). Internal forward models in the cerebellum: fMRI study on grip force and load force coupling. *Prog. Brain. Res.* **142,** 171–188.

Koslow, S. H. (2000). Should the neuroscience community make a paradigm shift to sharing primary data? *Nat. Neurosci.* **2,** 863–864.

Koslow, S. H. (2002). Opinion: Sharing primary data: A threat or asset to discovery? *Nat. Rev. Neurosci.* **3,** 311–313.

LaConte, S., Anderson, J., Muley, S., Ashe, J., Frutiger, S., Rehm, K., Hansen, L. K., Yacoub, E., Hu, X., Rottenberg, D., and Strother, S. (2003). The Evaluation of Preprocessing Choices in Single-Subject BOLD fMRI Using NPAIRS Performance Metrics. *Neuroimage* **18,** 10–27.

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y. F., Field, A. S., and Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *J. Cogn. Neurosci.* **14,** 420–429.

Legato, J., Knepper, M. A., Star, R. A., and Mejia, R. (2003). Database for renal collecting duct regulatory and transporter proteins. *Physiol. Genomics.* **13,** 179–181.

Leveroni, C. L., Seidenberg, M., Mayer, A. R., Mead, L. A., Binder, J. R., and Rao, S. M. (2000). Neural systems underlying the recognition of familiar and newly learned faces. *J. Neurosci.* **20,** 878–886.

Lindblom, A., Tannergard, P., Werelius, B., and Nordenskjold, M. (1993). Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat. Genet.* **5,** 279–282.

Liou, M., Su, H-R., Lee, J-D., and Cheng, P. E. (2003). Bridging functional MR images and scientific inference: Reproducibility maps. *J. Cog. Neurosci.* **15,** 934–945.

Lloyd, D. (2002). Functional MRI and the study of human consciousness. *J. Cogn. Neurosci.* **14,** 818–831.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412,** 150–157.

Lohmann, G., Muller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., Zysset, S., and von Cramon, D. Y. (2001). LIPSIA—a new software system for the evaluation of functional magnetic resonance images of the human brain. *Comput. Med. Imaging Graph* **25,** 449–457.

Lukic, A. S., Wernick, M. N., and Strother, S. C. (2002). An evaluation of methods for detecting brain activations from functional neuroimages. *Artif. Intell. Med.* **25,** 69–88.

Ma, B., Tromp, J., and Li, M. (2002). PatternHunter: Faster and more sensitive homology search. *Bioinformatics* **18,** 440–445.

Marenco, L., Nadkarni, P., Skoufos, E., Shepherd, G., and Miller, P. (1999). Neuronal database integration: The Senselab EAV data model. *Proc. AMIA Symp.* **10,** 102–106.

Martone, M., Gupta, A., Wong, M., Qian, X., Sosinsky, G., Ludascher, B., and Ellisman, M. (2002). A cell-centered database for electron tomographic data. *J. Struct. Biol.* **138,** 145.

Mavroudis, C., and Jacobs, J. P. (2000). Congenital heart surgery nomenclature and database project: Overview and minimum dataset. *Ann. Thorac. Surg.* **69,** S2–S17.

Mazziotta, J., Toga, A. W., Evans, A., Fox, P., Lancaster, J., Ziles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R., and Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos. Trans. R Soc. Lond. B Biol. Sci.* **356,** 1293–1322.

McKeown, M. J., and Sejnowski, T. J. (1998). Independent component analysis of fMRI data: Examining the assumptions. *Hum. Brain Mapp.* **6,** 368–372.

Mechelli, A., Friston, K. J., and Price, C. J. (2000). The effects of presentation rate during word and pseudoword reading: A comparison of PET and fMRI. *J. Cogn. Neurosci.* **12,** 145–156.

Mechelli, A., Gorno-Tempini, M. L., and Price, C. J. (2003a). Neuroimaging studies of word and pseudoword reading: Consistencies, inconsistencies, and limitations. *J. Cogn. Neurosci.* **15,** 260–271.

Mechelli, A., Henson, R. N., Price, C. J., and Friston, K. J. (2003b). Comparing event-related and epoch analysis in blocked design fMRI. *Neuroimage* **18,** 806–810.

Mechelli, A., Price, C., Noppeney, U., and Friston, K. (2003c). A dynamic causal modelling study on category effects: Bottom-up or top-down mediation? *J. Cogn. Neurosci.* **15,** 925–934.

Mirnics, K. (2001). Microarrays in brain research: The good, the bad and the ugly. *Nat. Rev. Neurosci.* **2,** 444–447.

Mitchell, T. (1997). "Machine Learning." McGraw-Hill, New York.

Mitchell, T. (1999). "Machine Learning and Data Mining." Communications of the ACM 42.

Mohamed, F. B., Tracy, J. I., Faro, S. H., Emperado, J., Koenigsberg, R., Pinus, A., and Tsai, F. Y. (2000). Investigation of alternating and continuous experimental task designs during single finger opposition fMRI: A comparative study. *J. Comput. Assist. Tomogr.* **24,** 935–941.

Noguchi, T., and Akiyama, Y. (2003). PDB-REPRDB: A database of representative protein chains from the Protein Data Bank (PDB) in 2003. *Nucleic Acids Res.* **31,** 492–493.

Nowinski, W. L., Belov, D., and Benabid, A. L. (2002). A community-centric internet portal for stereotactic and functional neurosurgery with a probabilistic functional atlas. *Stereotact. Funct. Neurosurg.* **79,** 1–12.

Ogata, H., Fujibuchi, W., Goto, S., and Kanehisa, M. (2000). A heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. *Nucleic Acids Res.* **28,** 4021–4028.

Oliver, D. E., Rubin, D. L., Stuart, J. M., Hewett, M., Klein, T. E., and Altman, R. B. (2002). Ontology development for a pharmacogenetics knowledge base. *Pac. Symp. Biocomput.* **6,** 5–76.

Peltomaki, P. T. (1994). Genetic basis of hereditary nonpolyposis colorectal carcinoma (HNPCC). *Ann. Med.* **26,** 215–219.

Penny, W. D., Stephan, K. E., Mechelli, A., and Friston, K. J. (2004). Comparing dynamic causal models. *Neuroimage* **22,** 1157–1172.

Persson, B. (2000). Bioinformatics in protein analysis. *Exs* **88,** 215–231.

Postle, B. R., Berger, J. S., Taich, A. M., and D'Esposito, M. (2000). Activity in human frontal cortex associated with spatial working memory and saccadic behavior. *J. Cogn. Neurosci.* **12,** 2–14.

Price, C. J., and Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends Cogn. Sci.* **6,** 416–421.

Rafalski, J. A., Hanafey, M., Miao, G. H., Ching, A., Lee, J. M., Dolan, M., and Tingey, S. (1998). New experimental and computational approaches to the analysis of gene expression. *Acta Biochim. Pol.* **45,** 929–934.

Raichle, M. E. (1975). Cerebral blood flow and metabolism. *Ciba. Found Symp.* 85–96.

Raichle, M. E. (2001a). Cognitive neuroscience. Bold insights. *Nature* **412,** 128–130.

Raichle, M. E. (2001b). Functional Neuroimaging: A historical and physiological perspective. *In* "Handbook of Functional Neuroimaging of Cognition" (R. Cabeza and A. Kingstone, Eds.), pp. 3–26. MIT Press, Cambridge, MA.

Raichle, M. E. (2003). Functional brain imaging and human brain function. *J. Neurosci.* **23,** 3959–3962.

Ramesar, R. S., Madden, M. V., Felix, R., Harocopos, C. J., Westbrook, C. A., Jones, G., Cruse, J. P., and Goldberg, P. A. (2000). Molecular genetics improves the management of hereditary non-polyposis colorectal cancer. *S Afr. Med. J.* **90,** 709–714.

Reidpath, D. D., and Allotey, P. A. (2001). Data sharing in medical research: An empirical investigation. *Bioethics* **15,** 125–134.

Rex, D. E., Ma, J. Q., and Toga, A. W. (2003). The LONI pipeline processing environment. *Neuroimage* **19,** 1033–1048.

Richard, A. M., and Williams, C. R. (2002). Distributed structure-searchable toxicity (DSSTox) public database network: A proposal. *Mutat. Res.* **499,** 27–52.

Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., and Shelton, J. N. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nat. Neurosci.* **6,** 1323–1328.

Roland, P., Svensson, G., Lindeberg, T., Risch, T., Baumann, P., Dehmel, A., Frederiksson, J., Halldorson, H., Forsberg, L., Young, J., and Zilles, K. (2001). A database generator for human brain imaging. *Trends Neurosci.* **24,** 562–564.

Rosen, B. R., Aronen, H. J., Kwong, K. K., Belliveau, J. W., Hamberg, L. M., and Fordham, J. A. (1993). Advances in clinical neuroimaging: Functional MR imaging techniques. *Radiographics* **13,** 889–896.

Rosen, B. R., Buckner, R. L., and Dale, A. (1998). Event-related functional MRI: Past, present, and future. *Proc. Nat. Acad. Sci. USA* **95,** 773–780.

Salli, E., Aronen, H. J., Savolainen, S., Korvenoja, A., and Visa, A. (2001). Contextual clustering for analysis of functional MRI data. *IEEE Trans. Med. Imaging* **20,** 403–414.

Savoy, R. L. (2001). History and future directions of human brain mapping and functional neuroimaging. *Acta Psychologica* **107,** 9–42.

Schutte, B. C., Mitros, J. P., Bartlett, J. A., Walters, J. D., Jia, H. P., Welsh, M. J., Casavant, T. L., and McCray, P. B., Jr. (2002). Discovery of five conserved beta -defensin gene clusters using a computational search strategy. *Proc. Natl. Acad. Sci. USA* **99,** 2129–2133.

Shepherd, G. M., Mirsky, J. S., Healy, M. D., Singer, M. S., Skoufos, E., Hines, M. S., Nadkarni, P. M., and Miller, P. L. (1998). The Human Brain Project: Neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data. *Trends Neurosci.* **21,** 460–468.

Smaglik, P. (2000). Internet gateway planned for neuroinformatics data. *Nature* **405,** 603.

Sokoloff, L. (1981). Relationships among local functional activity, energy metabolism, and blood flow in the central nervous system. *Fed. Proc.* **40,** 2311–2316.

Strother, S. C., Anderson, J., Hansen, L. K., Kjems, U., Kustra, R., Sidtis, J., Frutiger, S., Muley, S., LaConte, S., and Rottenberg, D. (2002). The quantitative evaluation of functional neuroimaging experiments: The NPAIRS data analysis framework. *Neuroimage* **15,** 747–771.

Teo, P. C., Sapiro, G., and Wandell, B. A. (1997). Creating connected representations of cortical gray matter for functional MRI visualization. *IEEE Trans. Med. Imaging* **16,** 852–863.

Thomas, G. (1994). Advances in the genetics and molecular biology of colorectal tumors. *Curr. Opin. Oncol.* **6,** 406–412.

Toga, A. W. (2002a). Imaging databases and neuroscience. *Neuroscientist* **8,** 423–436.

Toga, A. W. (2002b). The laboratory of neuro imaging: What it is, why it is, and how it came to be. *IEEE Trans. Med. Imaging* **21,** 1333–1343.

Tukey, J. W. (1977). Exploratory Data Analysis. Addison, Wesley.

Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., and Anderson, C. H. (2001a). An integrated software suite for surface-based analyses of cerebral cortex. *J. Am. Med. Inform. Assoc.* **8,** 443–459.

Van Essen, D. C., Lewis, J. W., Drury, H. A., Hadjikhani, N., Tootell, R. B., Bakircioglu, M., and Miller, M. I. (2001b). Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Res.* **41,** 1359–1378.

Van Horn, J. D. (2002). Maturing as a science: The new perspectives in fMRI research award. *J. Cogn. Neurosci.* **14,** 817.

Van Horn, J. D., Ellmore, T. M., Esposito, G., and Berman, K. F. (1998). Mapping voxel-based statistical power on parametric images. *Neuroimage* **7,** 97–107.

Van Horn, J. D., and Gazzaniga, M. S. (2002). Databasing fMRI studies—Toward a 'Discovery Science' of brain function. *Nat. Rev. Neurosci.* **3,** 314–318.

Van Horn, J. D., and Gazzaniga, M. S. (2004). Maximizing information content in shared neuroimaging studies of cognitive function. *In* "Databasing the Brain: From Data to Knowledge" (A. Subramanian, Ed.). John Wiley and Sons, New York.

Van Horn, J. D., Grethe, J. S., Kostelec, P., Woodward, J. B., Aslam, J. A., Rus, D., Rockmore, D., and Gazzaniga, M. S. (2001). The functional magnetic resonance imaging data center (fMRIDC): The challenges and rewards of large-scale databasing of neuroimaging studies. *Philos. Trans. R. Soc. Lond B. Biol. Sci.* **356,** 1323–1339.

Van Horn, J. D., Woodward, J. B., Simonds, G., Vance, B., Grethe, J. S., Montague, M., Aslam, J. A., Rus, D., Rockmore, D., and Gazzaniga, M. S. (2002). The fMRI data center: Software tools for neuroimaging data management, inspection, and sharing. *In* "A Practical Guide to Neuroscience Databases and Associated Tools" (R. Kotter, Ed.), pp. 221–235. Kluwer, Amsterdam.

van Stolk, R. U. (2002). Familial and inherited colorectal cancer: Endoscopic screening and surveillance. *Gastrointest Endosc. Clin. N. Am.* **12,** 111–133.

Velleman, P., and Hoaglin, D. (1981). The ABC's of EDA: Applications, basics, and computing of exploratory data analysis: Duxbury.

Wong, S. T., and Koslow, S. H. (2001). Human brain program research progress in bioinformatics/ neuroinformatics. *J. Am. Med. Inform. Assoc.* **8,** 103–104.

Woo, J. H., and Hathout, G. M. (2001). Systems analysis of functional magnetic resonance imaging data using a physiologic model of venous oxygenation. *J. Cereb. Blood. Flow. Metab.* **21,** 517–528.

Young, M. P., and Scannell, J. W. (2000). Brain structure-function relationships: Advances from neuroinformatics. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **355,** 3–6.

Further Reading

Druzgal, T. J., and D'Esposito, M. (2001). Activity in fusiform face area modulated as a function of working memory load. *Brain. Res. Cogn. Brain Res.* **10,** 355–364.