frontiers in
NEUROINFORMATICS

EDITORIAL
published: 25 September 2009
doi: 10.3389/neuro.11.031.2009

Neuroimaging workflow design and data-mining: a Frontiers in Neuroinformatics special issue

John Darrell Van Horn* and Arthur W. Toga

Laboratory of Neuro Imaging, Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
** Correspondence: jvanhorn@loni.ucla.edu*

The development of sophisticated neuroimaging data processing tools has been of major importance for distilling the large amount of information present in brain imaging data sets into useful and enlightening results. Neuroinformatics-based algorithms, in particular, have been instrumental in analyzing population level cortical anatomy, changes in BOLD activity, and, more recently, the rapid processing of diffusion weighted images (DTI/HARDI). Several notable examples include Statistical Parametric Mapping (Friston, 2006), FSL (Smith et al., 2004), FreeSurfer (surfer.nmr.mgh.harvard.edu), AFNI (Cox, 1996), and BrainVoyager (Goebel et al., 2006), among other analysis packages. The wide availability of neuroinformatics tools has helped to significantly spur growth in cognitive and clinical neuroscience, as well as permitted the efficient re-analysis of data contained in large-scale data archives (Kennedy and Haselgrove, 2006).

Within any of the aforementioned software packages it is possible to find the majority of individual steps needed for processing the most common types of brain imaging data. With these individual operations, accompanied by various inputs, parameters, and other options, investigators frequently link executable programs together as "scripts" or batch processes in which inputs are passed to one executable and the resulting outputs become the input to the next processing executable, and so on. In so doing, many laboratories have found it possible to create efficient yet flexible data processing streams to not only process data within modality but also between modalities. The notion of scientific workflows has now taken on its own formalism, moving from beyond custom-built scripts toward fully-fledged software environments with several available software platforms available to construct neuroimaging workflows, optimize their performance, and that take advantage of super-computing and grid infrastructures to expedite data processing throughput (Romano et al., 2005; Oinn et al., 2006; Van Horn et al., 2006; Ruping et al., 2007; Verdi et al., 2007). With a fully encompassing workflow platform it is also possible to break out of a "package-centric" view of neuroimage data processing and toward an informatics model that draws processing capabilities from across existing software suites as well as the incorporation of local informatics tools into heterogeneous analysis workflows. Such workflow descriptions, which themselves are often highly structured file formats describing the executable operations and their various processing choices, can serve to provide needed data provenance ensuring the fidelity of data reanalysis and replication (Mackenzie-Graham et al., 2008).

More than simply processing individual subject datasets or even the data from complete neuroimaging studies, the notion of workflows has permeated the next level of neuroimaging analysis beyond subject or study-based processing: that of data mining and meta-analysis. Data mining is a process of exploring data to identify potentially interesting patterns in the data that might not have been examined in the original research studies in question or perhaps were not detected by traditional statistical methods. These approaches to sifting through large archives of data to extract potentially useful patterns and relationships has been most evident in the genomic sciences although neuroimagers have explored these methods as well (Mitchell, 1999; Megalooikonomou et al., 2000; Wigle et al., 2001; Anderle et al., 2003). Meta-analysis, on the other hand, first gained the attention of the social sciences and related fields in the late 1970's and 1980's as a means to examine the study-specific and experimental factors that predicted reported effect sizes present in published studies (Glass et al., 1981; Rosenthal, 1984). The notion of performing an "analysis of analyses" to quantitatively critique, explore, and synthesize a literature has proved to be highly compelling and powerful. With the burgeoning growth of neuroimaging studies of brain structural differences between clinical populations and examinations of human cognition using PET and fMRI, the concept of meta-analysis soon found its way into the realm of brain imaging (Van Horn and McManus, 1992; Fox and Woldorff, 1994; Cabeza and Nyberg, 2000). Data mining and meta-analyses permit the exploration of not only the neural structure or patterns of cognitively-induced activity, but these analyses can also provide insights into those study factors that can predict the magnitude of reported effects. These approaches help to synthesize data from across studies, craft general trends in results from across studies, and quantify the effects of predictor variables obtained from the studies themselves that may influence the size and scale of differences.

Data processing workflow concepts have been an important element for meta-analyses and data mining, too, providing the basis for how sufficient summary metrics are obtained, combined with appropriate study meta-data, and then systematically compared and combined from across subjects and studies (**Figure 1**). Visualizing the relationships between subjects and study results has also been an important element for meta-analysis results and workflows are needed to provide graphical representations to still further neuroinformatics tools needed for dynamic and interactive visualization (Toga and Thompson, 2002; Van Essen, 2002; Van Essen and Dierker, 2007).

In this special issue of *Frontiers in Neuroinformatics*, we have invited several leading groups to provide articles focusing on the development of workflow technologies and perspectives for efficient neuroimage data processing and that help to permit subsequent meta-analysis of the results. Articles by Dinov et al., Ooi et al., Kenny et al., and Cheng et al. showcase recent developments in advanced workflow technologies for efficient processing of neuroimaging data. Contributions from Keator and colleagues discuss the use of high-performance computing capabilities upon which workflow environments have been specifically designed to take advantage of for rapid processing, while articles from Costafreda et al., Bockholt et al., Lohrey et al., and Laird et al. discuss the

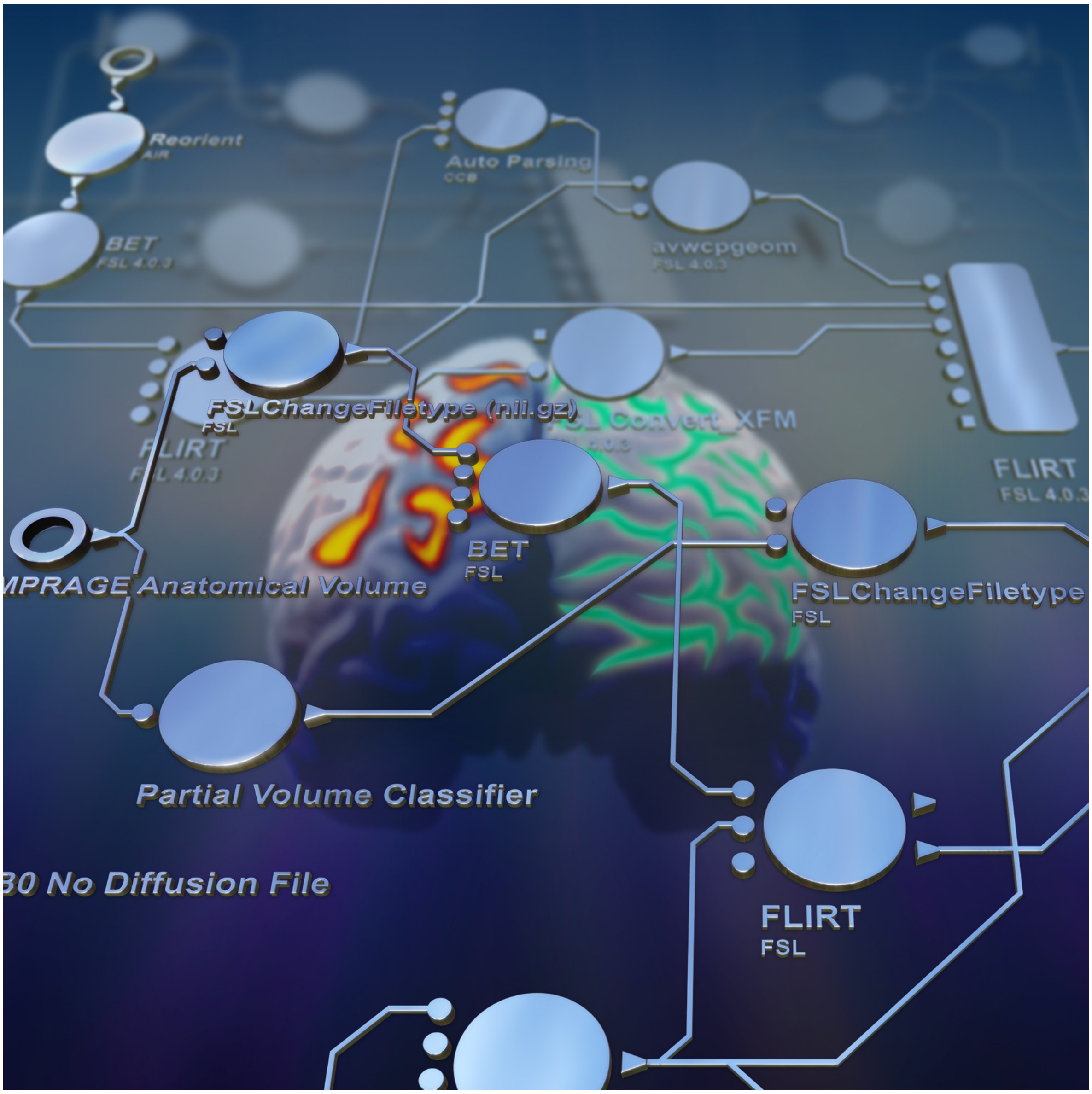

FIGURE 1 | Scientific workflows provide flexible platforms for multimodal neuroimage processing that facilitate high-throughput analysis of individual subjects as well as complete studies. These are also essential software and informatics frameworks for data mining and exploration, meta-analytic consideration of effects from across multiple studies, as well as providing efficient approaches for visualizing synthesized results and the functional/structural relationships that exist between brain imaging data sets.

development of data mining workflows and feature interesting examples of data synthesis. Finally, contributions from Nielsen and from Joshi et al. discuss the important role of visualization in data mining and the workflows necessary to inform novel informatics tools that focus on interactively exploring the relationships amongst large collections of brain data. The quality of these articles is exceptional and provides a broad overview at how workflow concepts have matured for the neuroimaging field, how they are now being used to expedite data mining, meta-analysis, and helping to provide the content needed for graphical data interaction.

Informatics as had a historical foothold in the data-rich field of neuroimaging. However, with *in vivo* datasets continually increasing in size, scope, and complexity, the continued development of efficient processing tools remains necessary to extract the maximal amount of useful information from them. Workflow technologies for data processing design, application, and execution link these tools into high-throughput processing pipelines. Their ongoing development can be expected to greatly enrich the ability of researchers to not only process newly obtained neuroimaging data but also to compare, contrast, and combine results from previous research studies via meta-analytic and data mining approaches and to visualize unique patterns present in neuroimaging results that could only be identified through large-scale informatics approaches.

REFERENCES

Anderle, P., Duval, M., Draghici, S., Kuklin, A., Littlejohn, T. G., Medrano, J. F., Vilanova, D., and Roberts, M. A. (2003). Gene expression databases and data mining. *Biotechniques* Suppl, 36–44.

Cabeza, R., and Nyberg, L. (2000). Imaging cognition II: an empirical review of 275 PET and fMRI studies. *J. Cogn. Neurosci.* 12, 1–47.

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

Fox, P. T., and Woldorff, M. G. (1994). Integrating human brain maps. *Curr. Opin. Neurobiol.* 4, 151–156.

Friston, K. J. (2006). Statistical Parametric Mapping. London, Academic Press.

Glass, G. V., McGaw, B., and Smith, M. L. (1981). Meta-Analysis in Social Research. Beverly Hills, Sage Publications.

Goebel, R., Esposito, F., and Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp.* 27, 392–401.

Kennedy, D. N., and Haselgrove, C. (2006). The internet analysis tools registry: a public resource for image analysis. *Neuroinformatics* 4, 263–270.

Mackenzie-Graham, A. J., Van Horn, J. D., Woods, R. P., Crawford, K. L., and Toga, A. W. (2008). Provenance in neuroimaging. *Neuroimage* 42, 178–195.

Megalooikonomou, V., Ford, J., Shen, L., Makedon, F., and Saykin, A. (2000). Data mining in brain imaging. *Stat. Methods Med. Res.* 9, 359–394.

Mitchell, T. (1999). Machine learning and data mining. *Commun. ACM* 42, 30–36.

Oinn, T., Greenwood, M., Addis, M., Alpdemir, M. N., Ferris, J., Glover, K., Goble, C., Goderis, A., Hull, D., Marvin, D., Li, P., Lord, P., Pocock, M. R., Senger, M., Stevens, R., Wipat, A., and Wroe, C. (2006). Taverna: lessons in creating a workflow environment for the life sciences. *Concurrency Comput. Pract. Exp.* 18, 1067–1100.

Romano, P., Marra, D., and Milanesi, L. (2005). Web services and workflow management for biological resources. *BMC Bioinformatics* 6(Suppl. 4), S24.

Rosenthal, R. (1984). Meta-Analytic Procedures for Social Research. Beverly Hills, Sage.

Ruping, S., Sfakianakis, S., and Tsiknakis, M. (2007). Extending workflow management for knowledge discovery in clinico-genomic data. *Stud. Health Technol. Inform.* 126, 184–193.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., and Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23(Suppl. 1), S208–S219.

Toga, A. W., and Thompson, P. M. (2002). New approaches in brain morphometry. *Am. J. Geriatr. Psychiatry* 10, 13–23.

Van Essen, D. C. (2002). Surface-based atlases of cerebellar cortex in the human, macaque, and mouse. *Ann. N. Y. Acad. Sci.* 978, 468–479.

Van Essen, D. C., and Dierker, D. L. (2007). Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron* 56, 209–225.

Van Horn, J. D., Dobson, J., Woodward, J., Wilde, M., Zhao, Y., Voeckler, J., and Foster, I. (2006). Grid-based computing and the future of neuroscience computation. In Methods in Mind, C. Senior, T. Russell, and M. S. Gazzaniga, eds (Cambridge, MIT Press), pp. 141–170.

Van Horn, J. D., and McManus, I. C. (1992). Ventricular enlargement in schizophrenia. A meta-analysis of studies of the ventricle:brain ratio (VBR). *Br. J. Psychiatry* 160, 687–697.

Verdi, K. K., Ellis, H. J., and Gryk, M. R. (2007). Conceptual-level workflow modeling of scientific experiments using NMR as a case study. *BMC Bioinformatics* 8, 31.

Wigle, D. A., Rossant, J., and Jurisica, I. (2001). Mining mouse microarray data. *Genome Biol.* 2, REVIEWS1019.

Received: 17 August 2009; published online: 25 September 2009

Citation: Front. Neuroinform. (2009) ***3****:31. doi: 10.3389/neuro.11.031.2009*