



Electronic data capture, representation, and applications for neuroimaging

Jessica A. Turner^{1*} and John Darrell Van Horn²

¹ Mind Research Network, University of New Mexico, Albuquerque, NM, USA

² Laboratory of Neuro Imaging, Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

*Correspondence: jturner@mm.org

INTRODUCTION

With growing emphasis on data sharing across neuroimaging studies in humans, both within and across institutions and studies, there is a growing awareness of the need for clear connections between the images and the other data about the subject, and between the images and the behavioral or physiological data that is relevant for analysis (Van Horn and Toga, 2009a). Means for electronic data capture, storage, organization, and levels of interactivity are required for the large and rich datasets which modern neuroimaging studies now generate and seek to analyze (Helmer et al., 2011). Beyond brain imaging volumes, researchers are including detailed meta-data on scan data types, cognitive protocols, phenomics, and graphical renderings as part of shareable dataset. Capturing these data using automated and semi-automated means represents a particular challenge for subsequent examination, mining, and visualization (Figure 1). The aim of this *Frontiers in Neuroinformatics* special topic was developed to sample from the current state of the art in automated and semi-automated data collection methods, data management, and their practical applications in neuroimaging research and related studies.

These articles, from experts in data capture, management, and visualization, provide an overview of the ways different research teams have addressed similar issues across the various stages of data capture in large-scale neuroimaging studies. Articles focus on how databases are populated, how to make sure that the information so gathered is accurate, and how to manage it so that it can be successfully communicated to others. Specifically, these articles address:

AUTOMATED AND SEMI-AUTOMATED DATA CAPTURE

Two of the papers deal directly with specific methods for data collection during the study. Voyvodic et al. (this issue) present a software package that has been applied in several studies both locally and across multiple sites. The CIGAL software makes a point of capturing behavioral and physiological data during experimental tasks in a way that is interpretable both during the experiment and afterward, and usable in single and multi-site neuroimaging or other behavioral studies. The timing of the various events during the experiment can be extracted from the resulting text files for storage in a database, translation into XML, or use in analysis pipelines.

Both the behavioral/physiological data capture of CIGAL and the clinical assessments capture of CARAT (Turner et al., 2010), make a point of representing the data in a way that is research-friendly and works with arbitrary data management systems. The CARAT package for collecting clinical measurements allows the user to connect to various databases to store the demographic and other data automatically, without need for transferring data from paper records.

PLATFORMS FOR DATA MANAGEMENT

Several papers in this issue present data management systems building on institutional level data capture, storage, protocol management, data review, and retrieval needs: COINS (Scott et al., this issue; building on MICIS, data warehousing for multiple studies and institutions), and LORIS (Das et al., this issue; building on their similar data warehouse experience), join the previously available LONI, XNAT (Marcus et al., 2007), and HID (Keator et al., 2008). Most of these systems have very similar specifications for similar types of data, the need for extensibility for new imaging techniques, the ability to pull imaging data from a DICOM receiver, sometimes with explicit interactions with other databases, and with different priorities for their end users. For instance, COINS facilitates cross-study and cross-institutional sharing, as well as user-based web portals for each study, LORIS has quality assurance (QA) measures which are quite sophisticated and protocol management/review as a core capability. LabIS (Prodanov et al., this issue) is an imaging data management system developed on animal studies, but the underlying needs to store subject information, imaging parameter values, and allow image viewing are in common with the systems developed for human research. LabIS has linked its schematic terms to standard vocabularies or ontologies, facilitating a clear data model for their uses.

DATA CURATION AND ANALYSIS

Imaging quality assurance (QA) measures and image browsing are a key step across all these data capture and managements systems. Data validation steps are included in CIGAL for behavioral data, in CARAT (client side) and LORIS (database side) for clinical assessments, and in COINS and LORIS for imaging protocols. Neuroimaging results (e.g., volumes, first order maps) are sometimes available.

DATA SHARING

A strength of the electronic data capture and management systems is to facilitate data sharing across researchers both within and between institutions. Poline et al. (this issue) describe the social and technological issues to large-scale data sharing in some detail. Neu et al. (this issue) provide an informative summary of the counterintuitive situation the neuroimaging research community is in – having both a lack of standards and a plethora of standards for storing even basic image information. COINS was designed to make this data sharing happen with no paperwork where possible; LORIS has been designed similarly, although LORIS is described as storing only de-identified data to make sharing very straightforward. COINS has various layers of security to protect subject confidentiality, but includes a summary data catalog or browser so uninvolved researchers can identify if there are data to which

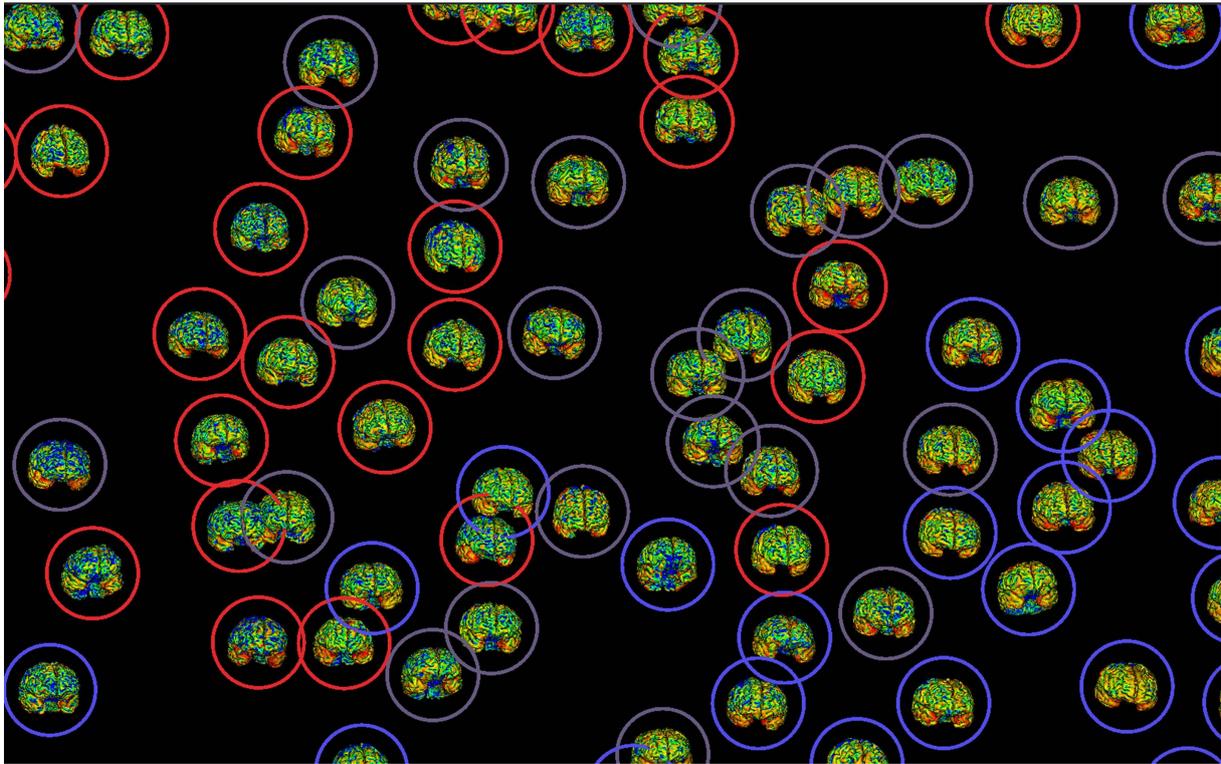


FIGURE 1 | Electronic data capture and efficient representation of neuroimaging enables comprehensive data mining and compelling visualization which, in turn, contributes to greater data sharing and openness in the brain sciences.

they wish to request access. These papers highlight some of the key challenges to electronic data capture in neuroimaging for easy and reliable data sharing.

INTERACTIVE DATA VISUALIZATION

While most of the neuroimaging data management systems include rudimentary image viewing on individual subjects, the INVIZIAN project (Bowman et al., this issue) draws imaging and meta-data from large-scale data archives, presenting it in a compelling graphical manner, allowing researchers to dynamically interact with, search, mine, and display complex representations from many hundreds of brain data sets simultaneously.

Semantic additions and the ability to understand what a dataset represents are addressed in part by the LabIS system (Prodanov et al., this issue), showing some of the strengths of those approaches. LabIs specifically focuses on the user needs for data annotation, linking imaging parameters, datasets, and their results to standardized terms from pre-existing ontologies.

NEXT STEPS

Neuroimaging has served as a focal point for data sharing for a number of years (Van Horn et al., 2004). However, as neuroimaging datasets increase in size, scope, and complexity, the development of efficient data capture methodologies, representational frameworks, and interactive technologies remains essential. Workflow technologies for data processing design and application will link these

tools into high-throughput processing pipelines, with consistent data capture, annotation, and provenance information available. Continued development of such approaches will enrich the ability of researchers to not only share newly obtained neuroimaging data, but also to combine and compare data via meta-analytic and data mining approaches, and to use large-scale integrative graphical tools to explore unique patterns only identifiable through these neuroinformatics approaches.

Linking captured and curated data to published research articles has been a long sought-after goal in neuroimaging (Van Horn et al., 2001; Van Horn and Toga, 2009b), integrating it with other biological data types (Ashish et al., 2010), and its workflow-based analysis under grid- or cloud-based computer systems (Van Horn et al., 2006; Keator et al., 2009; Dinov et al., 2010), thereby encouraging study data re-use (Van Horn and Ishai, 2007). Many issues remain to fully realize this goal including common challenges relevant to institutional review boards, and the need for standards for not only representing data and terminologies (Gupta et al., 2008) but permitting its integration with other schema (Van Horn and Ball, 2008; Gadde et al., 2011). Linking data and data models to semantic frameworks can already be seen in the LabIS system. Those links to ontologies or standard vocabularies allow the data repositories to be more easily understood by others (Bug et al., 2008), and ideally can capture knowledge not explicit in the database (Turner et al., 2010; Turner and Laird, 2011). The integration of databases with ontological and semantic frameworks promises to allow automated

reasoning over data and the results, accelerating progress in scientific research (to quote one of the reasons given by Poline et al. (this issue) for neuroimaging data sharing). The commonalities across the neuroimaging data management systems to date give reason for optimism that these standard vocabularies can be quickly developed

and accepted. Through this *Frontiers* special topic issue, we and the topic contributors seek to further encourage the prioritization of the role which neuroimaging data capture and sharing approaches play in our ability to easily exchange study data and efficiently document progress in the brain sciences.

REFERENCES

- Ashish, N., Ambite, J. L., Muslea, M., and Turner, J. A. (2010). Neuroscience data integration through mediation: an (F)BIRN Case Study. *Front. Neuroinform.* 4:118. doi: 10.3389/fninf.2010.00118
- Bug, W. J., Ascoli, G. A., Grethe, J. S., Gupta, A., Fennema-Notestine, C., Laird, A. R., Larson, S. D., Rubin, D., Shepherd, G. M., Turner, J. A., and Martone, M. E. (2008). The NIFSTD and BIRN Lex vocabularies: building comprehensive ontologies for neuroscience. *Neuroinformatics* 6, 175–194.
- Dinov, I., Lozev, K., Petrosyan, P., Liu, Z., Eggert, P., Pierce, J., Zamanyan, A., Chakrapani, S., Van Horn, J., Parker, D. S., Magsipoc, R., Leung, K., Gutman, B., Woods, R., and Toga, A. (2010). Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. *PLoS ONE* 5, e13070. doi: 10.1371/journal.pone.0013070
- Gadde, S., Aucoin, N., Grethe, J. S., Keator, D. B., Marcus, D. S., and Pieper, S. (2011). XCEDE: an extensible schema for biomedical data. *Neuroinformatics* 10, 19–32.
- Gupta, A., Bug, W., Marengo, L., Qian, X., Condit, C., Rangarajan, A., Muller, H. M., Miller, P. L., Sanders, B., Grethe, J. S., Astakhov, V., Shepherd, G., Sternberg, P. W., and Martone, M. E. (2008). Federated access to heterogeneous information resources in the Neuroscience Information Framework (NIF). *Neuroinformatics* 6, 205–217.
- Helmer, K. G., Ambite, J. L., Ames, J., Ananthakrishnan, R., Burns, G., Chervenak, A. L., Foster, I., Liming, L., Keator, D., Macciardi, F., Madduri, R., Navarro, J. P., Potkin, S., Rosen, B., Ruffins, S., Schuler, R., Turner, J. A., Toga, A., Williams, C., and Kesselman, C. (2011). Enabling collaborative research using the Biomedical Informatics Research Network (BIRN). *J. Am. Med. Inform. Assoc.* 18, 416–422.
- Keator, D. B., Grethe, J. S., Marcus, D., Ozyurt, B., Gadde, S., Murphy, S., Pieper, S., Greve, D., Notestine, R., Bockholt, H. J., Papadopoulos, P., Function, B., Morphometry, B., and Coordinating, B. (2008). A national human neuroimaging laboratory enabled by the Biomedical Informatics Research Network (BIRN). *IEEE Trans. Inf. Technol. Biomed.* 12, 162–172.
- Keator, D. B., Wei, D., Gadde, S., Bockholt, J., Grethe, J. S., Marcus, D., Aucoin, N., and Ozyurt, I. B. (2009). Derived data storage and exchange workflow for large-scale neuroimaging analyses on the BIRN grid. *Front. Neuroinform.* 3:30. doi: 10.3389/neuro.11.030.2009
- Marcus, D. S., Olsen, T. R., Ramaratnam, M., and Buckner, R. L. (2007). The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics* 5, 11–34.
- Turner, J. A., and Laird, A. R. (2011). The cognitive paradigm ontology: design and application. *Neuroinformatics* 10, 57–66.
- Turner, J. A., Mejino, J. L., Brinkley, J. F., Detwiler, L. T., Lee, H. J., Martone, M. E., and Rubin, D. L. (2010). Application of neuroanatomical ontologies for neuroimaging data annotation. *Front. Neuroinform.* 4:10. doi: 10.3389/fninf.2010.00010
- Van Horn, J. D., and Ball, C. A. (2008). Domain-specific data sharing in neuroscience: what do we have to learn from each other? *Neuroinformatics* 2, 117–121.
- 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*, eds C. Senior, T. Russell, and M. S. Gazzaniga (Cambridge: MIT Press), 141–170.
- Van Horn, J. D., Grafton, S. T., Rockmore, D., and Gazzaniga, M. S. (2004). Sharing neuroimaging studies of human cognition. *Nat. Neurosci.* 7, 473–481.
- 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., and Ishai, A. (2007). Mapping the human brain: new insights from fMRI data sharing. *Neuroinformatics* 5, 146–153.
- Van Horn, J. D., and Toga, A. W. (2009a). Is it time to re-prioritize neuroimaging databases and digital repositories? *Neuroimage* 47, 1720–1734.
- Van Horn, J. D., and Toga, A. W. (2009b). Multisite neuroimaging trials. *Curr. Opin. Neurol.* 22, 370–378.

Received: 02 April 2012; accepted: 11 April 2012; published online: 07 May 2012.
Citation: Turner JA and Van Horn JD (2012) Electronic data capture, representation, and applications for neuroimaging: a *Frontiers in Neuroinformatics Special Topic*. *Front. Neuroinform.* 6:16. doi: 10.3389/fninf.2012.00016
Copyright © 2012 Turner and Van Horn. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.