Neuroimage Data Sets: Rethinking Privacy Policies

Nakeisha Schimke and John Hale
Institute of Bioinformatics and Computational Biology
University of Tulsa, Tulsa, Oklahoma

Abstract

Data sharing has been promoted as a significant step in neuroimaging-based research for over a decade, yet the vision for widespread sharing has not been realized. Despite the availability of tools for deidentifying data and a few standout examples of data sharing, privacy concerns over the treatment of neuroimages as full face images have not been resolved. Continuing to share data, in large undertakings or ad hoc collaborations, without resolving the issue serves neither the subjects nor the researchers.

1 Introduction

Neuroimages are rich with data, from a subject's medical and personal history to image acquisition parameters. The depth of information invites data sharing in neuroimaging studies where the upfront costs of acquisition are high and subject enrollment tends to be low. Data sharing enables meta-analysis, data reuse, subject pooling, and study replication. As data crosses organizational boundaries, however, subject privacy must be addressed.

Consider the image itself–it may contain a condition, disease, diagnosis, or some benign but unusual feature. Or it may simply reveal that a particular scan was ordered. The image may contain unique features that reveal the subject's identity. The face is more closely associated with one's identity than other features, and neuroimages are unique in containing facial features. The possibility for reidentification has been suggested within the neuroscience community [9], but there is still no consensus on the severity of the issue.

The neuroscience community, if it wishes to continue sharing data, must reach an agreement on the reidentification issue. Avoiding the problem (or lack thereof) benefits neither subjects nor the research institutions. If structural neuroimages are equivalent to full face images, distributing them as deidentified data sets carries the responsibility of informing subjects of potential harm. While all data carries some risk of reidentification, subjects should be informed that their likeness is being transmitted even if their name is not. The nature of PHI and privacy in general are gathering more widespread attention, and financial penalties are being levied where the law is clearly violated. Ignoring the issue may prove catastrophic to organizations in an evolving legal and regulatory environment that continues to shape the definition of medical privacy. Worse still, the ad hoc and patchwork implementation of deidentification policies and practices could give rise to a convoluted sensitivity to the privacy threat where little or none may exist.

2 Deidentification tools

The number of existing privacy solutions far exceeds their practical deployment. There are several options for deidentifying neuroimages at the voxel level and even more at the metadata level. Metadata is relatively easy to remove by deleting or replacing offending values, and a wide range of standalone and integrated tools are available. The XNAT Redaction Toolkit is an integrated redaction workflow for XNAT, a neuroimaging collaboration framework [1]. It removes metadata from neuroimages and maintains persistent pseudonymous identifiers in a secure database to manage subjects across studies and sessions.

For the image data itself, there are two approaches: defacing and skull stripping. Skull stripping is a process to remove extraneous non-brain tissue from a neuroimage, routinely performed as part of the analysis workflow. The result is a brain-only volume without facial features. Skull stripping is effective, but the potential to discard valuable brain tissue makes it less attractive. There is also high dependence on parameter selection, and some techniques may favor a specific anatomical region. Skull stripping methods are detailed and compared in [4].

Defacing approaches leave behind extraneous brain tissue and remove only non-brain facial features, making it a more favorable solution for secondary use. MRI Defacer [3] uses a face atlas to identify and remove voxels with zero probability of containing brain tissue and non-zero probability of being face, effectively eroding facial features. Quickshear [6, 7] defacing is an alternative solution that eliminates the need for a manually labeled face atlas, relying instead on brain masking techniques to identify a plane that divides the neuroimage into two volumes, a face and a brain volume. The voxels of the "face" side are sheared off, removing identifiable facial features and leaving the brain volume intact.

3 Privacy in practice

A plethora of existing tools implement various levels of deidentification, but the use of defacing tools in practice is sparse. Data sharing has promised to transform

neuroscience research, but after twelve years of discussion, neuroscientists are still pondering the same issues. Though data sharing has not occurred on the scale envisioned by pioneers in the field, there are a few standout examples.

The fMRI Data Center (fMRIDC) effort collected data from 2000 to 2006. Uploads were mandatory for papers published in the *Journal of Cognitive Neuroscience*. Due to a lack of funding, fMRIDC stopped accepting new submissions in 2006, though it is still considered a successful pioneer in neuroimage data sharing. It demonstrated the usefulness of collaborative efforts and the added value of responsible data re-use. Its official privacy policy requested that uploads be deidentified, and fMRIDC would remove metadata and skull strip as necessary [9, 10].

The LONI Image Data Archive (IDA) is a central repository for neuroimaging and associated data. It provides the underlying informatics core for the Alzheimer's Disease Neuroimaging Initiative (ADNI), a collaborative undertaking that gathers data from multiple sites and distributes them to researchers around the world [5, 8]. The IDA suite of tools includes the LONI Deidentification Debablet for removing metadata, and only deidentified data sets are released. ADNI administrators track usage of and control access to the data, and researchers must agree to make no attempts to reidentify images.

The Biomedical Informatics Research Network (BIRN) provides the BIRN Deidentification Upload and Pipeline (BIRNDUP) tool to remove metadata and deidentify image data using MRI Defacer [2]. BIRNDUP is still available but further development has been halted until the BIRN community can reach a consensus on the need for image deidentification.

There are a number of economic, sociological, and academic obstacles that hamper collaboration on a large scale, but the privacy issue is one that can be easily remedied. Large scale data sharing efforts have strict access controls and policies in place, and the large number of available tools suggests that collaboration does occur. While ad hoc collaboration is still subject to legislative and IRB compliance, there are no overarching privacy policies for the neuroimage data itself.

4 Progress through policy

With regards to deidentification of the image data, there is no consensus, and the inability to come to an agreement has halted some deidentification efforts like BIRNDUP. The prevailing solution prohibits reidentification through data use agreements–a policy that insulates the originating institution from liability but does little to shield subjects from actual harm. Such agreements are sufficient for legitimate uses, but providers cannot expect adversarial users to adhere to these policies. Yet, data is being distributed as deidentified that may contain a full face image. It is reasonable that subjects should be informed that their likeness is being transmitted.

The resources of a combined effort to determine if the issue of reidentification exists, regardless of the outcome, serve to both inform subjects and safeguard institutions from financial and legal penalties.

Acknowledgments

References

[1] A. Barclay, N. Schimke, and J. Hale. Redacting PHI in Neurological Images using XNAT. Poster at USENIX Security Symposium 2010, Washington, DC, August 2010.

[2] MBIRN:BIRNDUP:Distribution, NAMIC Wiki. `http://www.na-mic.org/Wiki/index.php/MBIRN:BIRNDUP:Distribution`, 2008. Last accessed March 19, 2012.

[3] A. Bischoff-Grethe et al. A technique for the deidentification of structural brain MR images. *Human Brain Mapping*, 28:892–903, 2007.

[4] Christine Fennema-Notestine et al. Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: Effects of diagnosis, bias correction, and slice location. *Human Brain Mapping*, 27:99–113, 2006.

[5] S. G. Mueller et al. Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Neuroimaging Clincs of North America*, 2005.

[6] N. Schimke, M. Kuehler, and J. Hale. Preserving Privacy in Structural Neuroimages. In Y. Li, editor, *Data and Applications Security and Privacy XXV*, volume 6818 of *Lecture Notes in Computer Science*, pages 301–308. Springer Berlin / Heidelberg, 2011.

[7] N. Schimke, M. Kuehler, and J. Hale. Quickshear Defacing for Neuroimages. 2nd USENIX Workshop on Health Security and Privacy, San Francisco, California, 2011.

[8] Arthur W. Toga, Karen L. Crawford, and the Alzheimer's Disease Neuroimaging Initiative. The informatics core of the Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's & Dementia*, 6(3):247–256, 2010.

[9] John D. Van Horn et al. The Functional Magnetic Resonance Imaging Data Center (fMRIDC): the challenges and rewards of large-scale databasing of neuroimaging studies. *Philosophical Transactions: Biological Sciences*, 356:1323–1339, 2001.

[10] John D. Van Horn and A. W. Toga. Is it time to re-prioritize neuroimaging databases and digital repositories? *Neuroimage*, 47:1720–1734, 2009.