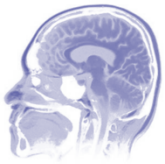
Redacting PHI in Neurological Images using XNAT

Alex Barclay – Computer Scientist[1] Ph.D. Candidate[2]
Nakeisha Schimke – Ph.D. Candidate[2]
Dr. John Hale, Ph.D. – Professor of Computer Science[2]

1 – Laureate Institute for Brain Research, Tulsa OK, USA
2 – University of Tulsa, Tulsa OK, USA
Institute of Bioinformatics and Computational Biology



The Problem

- Large-scale studies have huge amounts of data (1PB/3Yrs)
- Data shared must ensure privacy of subject (HIPAA)
- Inter-organizational collaboration must be easy
- Data exists as multiple abstractions, and simply removing it from a single layer is insufficient
- Tool is needed to specifically redact entire data stack of PHI and share data

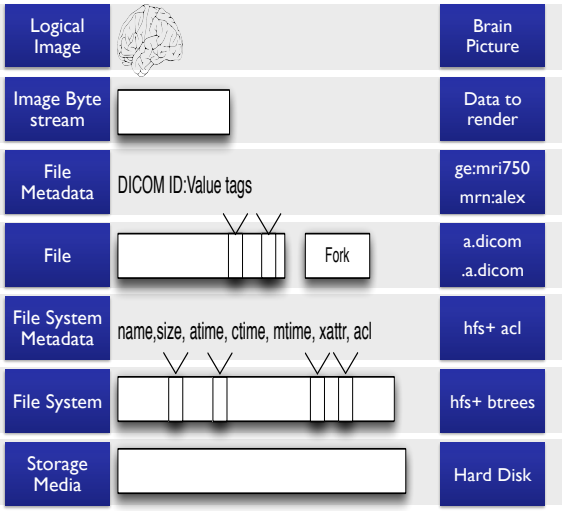
Why redact?

XNAT Background

- Neuroimaging data management platform www.xnat.org
- Combination relational database and file data store
- Remote data processing/execution pipepine
- Jakarta Turbine Web Application Framework / REST Interface
- Java with XML configuration files
- PostgreSQL database and Unix filesystem file store

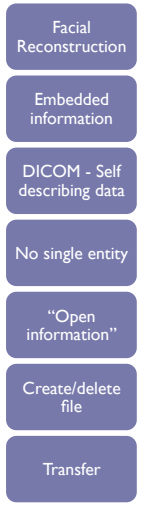
Data Stack

- Logical (not architectural or physical) break down of different storage and display components
- Bottom up approach to understand what contains PHI



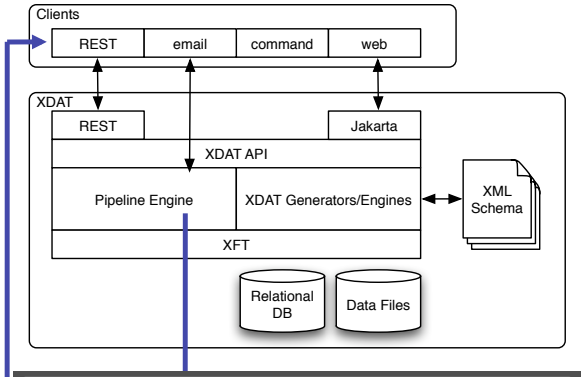
Issues

- Neuroscience
- Operating Sys
- Forensics



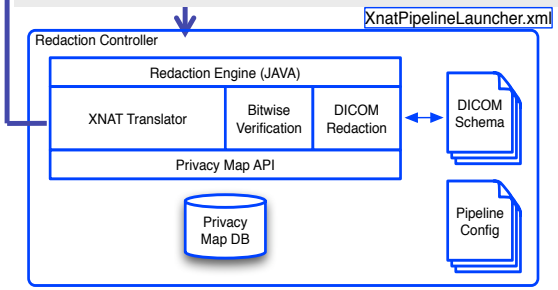
XNAT Architecture

- Developed as separate units:
  - Clients (Turbine/REST/E-mail)
  - Core (XML configured/Java middleware)
  - Pipeline (XML configured/implementation independent)

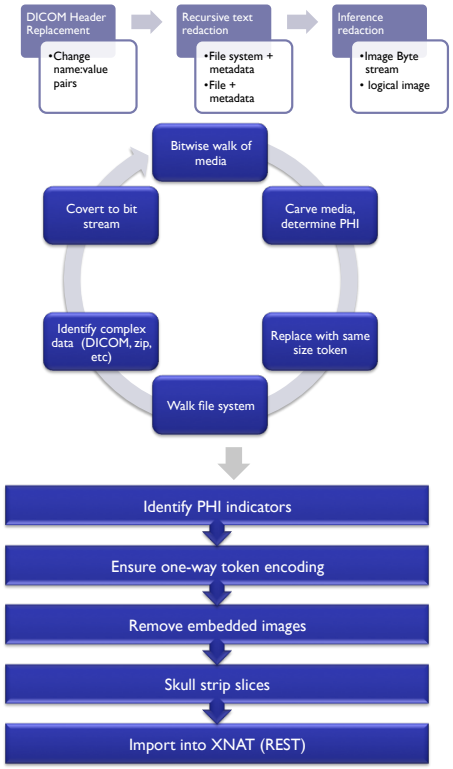


Modified/Added Architecture

- Builds upon XNAT pipeline interface (XML/Java/Python)
- Input: subjects with full scan data (DICOM)
- Output: mapped subject identifiers coupled to redacted images



PHI Redaction Engine



Discussion + Future Work

- Architecture and implementation of issues presented in poster at USENIX Security 09
- Comprehensive redaction is combination of recursive redaction at the block and file layer, with additional techniques to find and reduce inferred data such as DICOM
- Based upon body of work and code for legal production
- Project is started with goal to finish by August 2011

References:
Arkfeld, M. R. (2005), *Electronic Discovery and Evidence, Law Partner Publishing, L.L.C.*
Bischoff-Grethe, A., et al. (2007), *A technique for the Deidentification of Structural Brain MR Images. Human Brain Mapping, 28:892-903.*
Carrier, B. (2005), *File System Forensic Analysis, Addison-Wesley Professional.*
G. Manes, L. Watson, E. Downing, A. Barclay, D. Greer, J. Hale, *A Framework for Redacting Digital Information from Electronic Devices,* Proceedings of the 8th Annual IEEE SMC Information Assurance Workshop, West Point, New York, June 20-22, 2007.
Marcus, D. S, Oslen, 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**(1), 11-34.

Acknowledgments:
We gratefully acknowledge support from the William K. Warren Foundation, The National Institute of Health (1R03EB011961-01).