Appendix A

Regular Expressions Used to Remove PHI

Category	**Expression**
Social Security numbers	[0-9]{3}\-?[0-9]{2}\-?[0-9]{4}|[0-9]{9}
Phone numbers	\(([2-9]\d\d)\) ?[2-9]\d\d- ?\d\d\d\d|[2-9]\d\d\-? ?[2-9]\d\d-? ?[0-9]{4}|[2-9]\d\d\ ?[0-9]{4}|[2-9]\d\d\-? ?[0-9]{4}
Dates of birth	\bDOB: ?[A-Z]{3} ?[^A-Z]{7}|\bDATE OF BIRTH: ?[A-Z]{3} ?[^A-Z]{7}
Names	\b[A-Z]+,\s?[A-Z]+ \s?[B-H J-Z]\.?\b|\b[A-Z]+ \s?[B-H J-Z]\.? \s?[A-Z]+ \b(MR\b|MRS|MS)\. ?\s*[\w]*\b
Patient name	Patient Name \(Last, First, Middle\):\s*[\w.]+,?(\s*)?([\w.]+)(\s*)?([\w.]+)? PATIENT NAME:\s*[\w.]+(\s*)?([\w.]+),?(\s*)?([\w.]+)? Patient:\s*[\w.]+,?(\s*)?([\w.]+) PT:\s*[\w]+
Doctor names	\bRequesting Clinician:\s*[\w.]+,?(\s*)?([\w.]+)(\s*)?([\w.]+)? \bAuthor:\s*[\w.]+,?(\s*)?([\w.]+)(\s*)?([\w.]+)? \bSigned by:\s*[\w.]+,?(\s*)?([\w.]+)(\s*)?([\w.]+)? \bReading MD:\s*[\w.]+,?(\s*)?([\w.]+)(\s*)?([\w.]+)? \bTranscriber:\s*[\w.]+,?(\s*)?([\w.]+)(\s*)?([\w.]+)? \bExpected Cosigner:\s*[\w.]+,?(\s*)?([\w.]+)(\s*)?([\w.]+)? \bDR\.?:?\s*\b[\w]*\b (AUTHOR:).*(\s*)?(EXP) (COSIGNER:)\s*?[A-Z]+,?(\s*)?[A-Z]+ ([A-Z]+)? \/es\/\s*?[A-Z]+\s*[A-Z]+.?-?\s*([A-Z]+)?-?([A-Z]+)?/\/es\ \b[A-Z]+\b,\s*?([A-Z]+\.?)\s*([A-Z]+\.?)? ?,?\s*M\.?D\.?|(\b[A-Z]+\b)\s*[A-Z]+\s*([A-Z]+)? ?,?\s*M\.?D\.? \bPHYSICIAN:\s*?[A-Z]+,?\s*?[A-Z]+(\s*[A-Z]+)? (PROVIDER:)\s*?[A-Z]+,?\s*?[A-Z]+.?(\s*[A-Z]+)?
Facilities	(FACILITY:)\s*[A-Z]+(\s*[A-Z]+)?(\s*[A-Z]+)?(\s*[A-Z]+)?
Addresses	(ADDRESS:).*(\s*)?(CITY:)
ID numbers	\b\d{6,}\b \b[A-Z]+_?-?\d{3,}\b \b[A-Z][A-Z]*[0-9]+[\w]*\b
Dates	(((JAN)(UARY)?|(FEB)(UARY)?|(MAR)(CH)?|(APR)(IL)?|(MAY)|(JUN)E?|(JUL)Y?|(AUG)(UST)?|(SEP)(TEMBER)?|(OCT)(OBER)?|(NOV)(EMBER)?|(DEC)(EMBER)?) \s?[0-9][0-9]?,?\s?[0-9]{2,4}|[0-9]{1,2}\/[0-9]{1,2}\/[0-9]{2,4}|\b\d\d-?\d\d-?[12][90]\d\d\b|\b\d\d\/\d\d\/[12][90]\d\d\b) (\s\d\d:\d\d)? ([0-9]{1,2}\/[0-9]{1,2}\/[0-9]{2,4}|[0-9]{1,2}-[0-9]{1,2}-[0-9]{2,4}|\d\d-?\d\d-?[12][90]\d\d|\d\d\/\d\d\/[12][90]\d\d) (\s\d\d:\d\d)? \d\/[0-9]{1,2}\/[0-9]{2,4}|\d-[0-9]{1,2}-[0-9]{2,4} \d\/[0-9]{1,2}\/[0-9]{2,4} \d\d?\/[12][90]\d\d|\d\d?-[12][90]\d\d|\d\d?\/0\d|\d\d?-0\d \d\d?\s((JAN)(UARY)?|(FEB)(UARY)?|(MAR)(CH)?|(APR)(IL)?|(MAY)|(JUN)E?|(JUL)Y?|(AUG)(UST)?|(SEP)(TEMBER)?|(OCT)(TOBER)?|(NOV)(EMBER)?|(DEC)(EMBER)?)\s?\d\d(\d\d)?