Supplementary Table 1*. Sensitive Information Categories and Relevant Regular Expressions*

<table>
<tr><th>Sensitive Information</th><th>Subcategory</th><th>Regular expression patterns</th><th>Explanations</th></tr>
<tr><td rowspan="3"><b>(A) Names</b></td><td>Firstname, lastame</td><td>^[A-Z]\'?[-a-zA-Z]+$</td><td>match names, A'Bsfs, Absssfs, A-Bsfsfs</td></tr>
<tr><td>Title</td><td>(Dr\.|Mr\.|Mrs\.|Ms\.|Miss|Sir|Madam)\s
(([A-Z]\'?[A-Z]?[\-a-z]+(\s[A-Z]\'?[A-Z]?[\-a-z]+)*)
)</td><td>match salutation</td></tr>
<tr><td>Middle name</td><td>**PHI**,? (([A-CE-LN-Z][Rr]?|[DM])\.?) | (([A-CE-LN-Z][Rr]?|[DM])\.?),? **PHI**</td><td>match middle initial</td></tr>
<tr><td rowspan="2"><b>(B) Geographic subdivisions</b></td><td>Postal code</td><td>\b(\d{5}(-\d{4})?)\b</td><td>match postal code</td></tr>
<tr><td>Address</td><td>address_indictor = ['street', 'avenue', 'road', 'boulevard', 'drive', 'trail', 'way', 'lane', 'ave', 'blvd', 'st', 'rd', 'trl', 'wy', 'ln', 'court', 'ct', 'place', 'plc', 'terrace', 'ter', 'highway', 'freeway', 'autoroute', 'autobahn', 'expressway', 'autostrasse', 'autostrada', 'byway', 'auto-estrada', 'motorway', 'avenue', 'boulevard', 'road', 'street', 'alley', 'bay', 'drive', 'gardens', 'gate', 'grove', 'heights', 'highlands', 'lane', 'mews', 'pathway', 'terrace', 'trail', 'vale', 'view', 'walk', 'way', 'close', 'court', 'place', 'cove', 'circle', 'crescent', 'square', 'loop', 'hill', 'causeway', 'canyon', 'parkway', 'esplanade', 'approach', 'parade', 'park', 'plaza', 'promenade', 'quay', 'bypass'
'dr.', 'ave.', 'rd.', 'st.', 'blvd.','pkwy.','city']</td><td>address keywords</td></tr>
<tr><td><b>(C) Dates</b></td><td>General date</td><td>\b(.*?(?=\b(\d{1,2}[-./\s]\d{1,2}[-./\s]\d{2}|\d{1,2}[-./\s]\d{1,2}[-./\s]\d{4}|\d{2}[-./\s]\d{1,2}[-./\s]\d{1,2}|\d{4}[-./\s]\d{1,2}[-./\s]\d{1,2})\b))\b

month_name = Jan(uary)?|Feb(ruary)?|Mar(ch)?|Apr(il)?|May|Jun(e)?|Jul(y)?|Aug(ust)?|Sep(tember)?|Oct(ober)?|Nov(ember)?|Dec(ember)?

\b(</td><td>Date in various formats.
YYYY/MM-YYYY/MM
MM/YYYY-MM/YYYY
MM/YY-MM/YY
MM/YY-MM/YY
MM/YYYY-MM/YYYY
MM/DD-MM/DD
DD/MM-DD/MM
DD/MM-DD/MM
MM/DD/YY
MM/DD/YYYY</td></tr>
</table>

<table>
<tr><td></td><td></td><td>\d{4}[\-/](0?[1-9]|1[0-2]|"""+month_name+r""")\-\d{4}[\-/](0?[1-9]|1[0-2]|"""+month_name+r""") # YYYY/MM-YYYY/MM
|(0?[1-9]|1[0-2]|"""+month_name+r""")[\-/]\d{4}\-(0?[1-9]|1[0-2]|"""+month_name+r""")[\-/]\d{4} # MM/YYYY-MM/YYYY
|(0?[1-9]|1[0-2]|"""+month_name+r""")/\d{2}\-(0?[1-9]|1[0-2]|"""+month_name+r""")/\d{2} # MM/YY-MM/YY
|(0?[1-9]|1[0-2]|"""+month_name+r""")/\d{2}\-(0?[1-9]|1[0-2]|"""+month_name+r""")/\d{4} # MM/YYYY-MM/YYYY
|(0?[1-9]|1[0-2]|"""+month_name+r""")/([1-2][0-9]|3[0-1]|0?[1-9])\-(0?[1-9]|1[0-2]|"""+month_name+r""")/([1-2][0-9]|3[0-1]|0?[1-9]) #MM/DD-MM/DD
|([1-2][0-9]|3[0-1]|0?[1-9])/(0?[1-9]|1[0-2]|"""+month_name+r""")\-([1-2][0-9]|3[0-1]|0?[1-9])/(0?[1-9]|1[0-2]|"""+month_name+r""") #DD/MM-DD/MM
|(0?[1-9]|1[0-2]|"""+month_name+r""")[\-/\s]([1-2][0-9]|3[0-1]|0?[1-9])[\-/\s]\d{2} # MM/DD/YY
|(0?[1-9]|1[0-2]|"""+month_name+r""")[\-/\s]([1-2][0-9]|3[0-1]|0?[1-9])[\-/\s]\d{4} # MM/DD/YYYY
|([1-2][0-9]|3[0-1]|0?[1-9])[\-/\s](0?[1-9]|1[0-2]|"""+month_name+r""")[\-/\s]\d{2} # DD/MM/YY
|([1-2][0-9]|3[0-1]|0?[1-9])[\-/\s](0?[1-9]|1[0-2]|"""+month_name+r""")[\-/\s]\d{4} # DD/MM/YYYY
|\d{2}[\-./\s](0?[1-9]|1[0-2]|"""+month_name+r""")[\-\./\s]([1-2][0-9]|3[0-1]|0?[1-9])   # YY/MM/DD
|\d{4}[\-./\s](0?[1-9]|1[0-2]|"""+month_name+r""")[\-\./\s]([1-2][0-9]|3[0-1]|0?[1-9])   # YYYY/MM/DD
|\d{4}[\-/](0?[1-9]|1[0-2]|"""+month_name+r""") # YYYY/MM
|(0?[1-9]|1[0-2]|"""+month_name+r""")[\-/]\d{4} # MM/YYYY</td><td>DD/MM/YY
DD/MM/YYYY
YY/MM/DD
YYYY/MM/DD
MM/YYYY
MM/YY
MM/YYYY
MM/DD
DD/MM</td></tr>
</table>

		|(0?[1-9]|1[0-2]|"""+month_name+r""")/\d{2} # MM/YY |(0?[1-9]|1[0-2]|"""+month_name+r""")/\d{2} # MM/YYYY |(0?[1-9]|1[0-2]|"""+month_name+r""")/([1-2][0-9]|3[0-1]|0?[1-9]) #MM/DD |([1-2][0-9]|3[0-1]|0?[1-9])/(0?[1-9]|1[0-2]|"""+month_name+r""") #DD/MM)\b	
	DOB	\b(.*?(?=\b(\d{1,2}[-./\s]\d{1,2}[-./\s]\d{2} # X/X/XX |\d{1,2}[-./\s]\d{1,2}[-./\s]\d{4} # XX/XX/XXXX |\d{2}[-./\s]\d{1,2}[-./\s]\d{1,2} # xx/xx/xx |\d{4}[-./\s]\d{1,2}[-./\s]\d{1,2} # xxxx/xx/xx)\b))\b	Match date of birth
	Age	\b(age|year[s-]?\s?old|y.o[.]?)\b	Match age
(D)Telephone/ FAX numbers, (I) Health plan beneficiary numbers, Account numbers, Medical record numbers		\b((\d[\(\)\-\']?\s?){6}([\(\)\-\']?\d)+ |(\d[\(\)\-.\']?){7}([\(\)\-.\']?\d)+ # test)\b \b(\d{5}[A-Z0-9]*)\b \b([A-Z0-9\-/]{6}[A-Z0-9\-/]*)\b	Phone/fax/account number/MRN, etc. in the format of SSN/PHONE/FAX XXX-XX-XXXX, XXX-XXX-XXXX, XXX-XXXXXXXX, etc.
(F) Email addresses		\b([a-zA-Z0-9_.+-@\"]+@[a-zA-Z0-9-\:\]\[]+[a-zA-Z0-9-.]*)\b	e.g., xxx.xxx@xxxx.xxx
(G) Social security numbers		\b((\d[\(\)\-\']?\s?){6}([\(\)\-\']?\d)+|(\d[\(\)\-.\']?){7}([\(\)\-.\']?\d)+)\b	e.g., xxx-xx-xxxx, etc.
(N) Web URLs		\b((http[s]?://)?([a-zA-Z0-9$-_@.&+:!*\(\),])*[\.\/]([a-zA-Z0-9$-_@.&+:\!*\(\),])*)\b	e.g., http://xxx.xxx.xxx/xxx.xxx

<table>
<tr><td><b>(O) Internet Protocol (IP) addresses</b></td><td></td><td>^(?:(?:25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)\.){3}(?:25[0-5]|2[0-4][0-9]|[01]?[0-9][0-9]?)$</td><td>e.g., xxx.xxx.xxx.xxx</td></tr>
<tr><td><b>(R) Other unique identifiers</b></td><td>Race</td><td>“(White/Caucasian|White|Caucasian|American|Indian|Alaska|Native|Asian|Black|African|Native|Hawaiian|Pacific|Islander)”</td><td>Match race</td></tr>
<tr><td></td><td>Gender</td><td>"(Male|Female)"</td><td>Match gender</td></tr>
<tr><td></td><td>Ethnicity</td><td>"(Hispanic|Latino)"</td><td>Match ethnicity</td></tr>
</table>