Security Language
Previously Approved

Social Media—Twitter

Loss of confidentiality has to be taken into consideration, although this would likely not be serious for the subjects since the primary data collected has already been consciously and actively made public by the enrolled subject on her social network. By joining the online application, subjects grant access to personal and identifying information associated with their social network accounts. However, we will only store and analyze published information pertaining to their social network in a confidential way with unique and confidential patient identifiers.

Twitter *user handles* are the publicly available user information. All personal identifiable information contained in the tweets (namely the Twitter *user handles* and user information about tweet authors, followers, and followees) will be redacted immediately and automatically assigned a unique confidential identifier. Confidential identifiers will allow us to characterize the social connections across the network, without the need to maintain private identifiable information. We will also immediately assign confidential identifiers to any third parties identified through private tweets.

Collecting posted information within the social network is key to effectively study the participants' social network. Since this research presents minimal risks, it will not adversely affect the rights and welfare of the online community of the enrolled subject. In the event that a study participant engages in any private communication (ie NOT a part of the public data domain) with a third party, we will only be able to monitor and store the study participant's data. In other words, we will only have access to one side of any private conversation: the consented participant's side. We will not have access to any third party private information. Twitter handle will be immediately turned into a unique ID and the content of the tweets will be redacted to replace any direct reference to another user with the unique ID assigned to that user. Additionally, retweets from a third party who has not consented to participate in this study will not be available for researchers. Moreover, after classifying the tweets we will filter them to attempt to remove PII. Given the heterogeneous way that PII can be expressed in tweets, note that there is no guarantee that all PII information will be conceived. Stored confidential data will be accessible only via password-protected secure login. Strong passwords will be chosen to protect confidentiality. Only the PIs and designated research assistants will be allowed to access the data, and no identifying information will ever leave the secure environment. The risk to users is unlikely to be significantly different from the risk they already bear by having their data stored on the publicly available Twitter servers.

Should unanticipated events involving risks to subjects or others occur, the reporting of unanticipated problems and adverse events will occur in a timely manner. The unanticipated event will be reported to the IRB and appropriate institutional officials within 2 weeks of the investigator becoming aware of the event.