RESEARCH SHORTS #2

Collecting Social Media Data for Qualitative Research

June 2015

Social media activity among contemporary Western youth is almost universal. For most youth it is inextricably woven into their daily activities and serves as a primary means of both expression and communication. Capturing and analyzing this online activity can open new windows for researchers into the everyday lives and concerns of youth. Collecting social media data, however, can be ethically and procedurally challenging. This paper outlines the protocols utilized by the *Young Lives Research Lab* for this data collection, and our recommendations to other researchers based on the efficacy of our past efforts.

For a longitudinal project examining the impacts of digital media on the lives of youth over time, the team endeavored to collect participants' social media activity over a period of 3 months to complement data collected via interviews and other qualitative tools. In order to mitigate the vast amount of data that could be collected, we utilized pilot interview discussions with youth to establish the 5 most popular social media platforms at the time; these were identified as: Facebook, Twitter, Tumblr, Instagram and Pinterest (see sidebar). As the popularity of these platforms can change quickly, the most common platforms being used by youth should be re-evaluated at the start of any research project.

A Social Media Primer

Facebook: Members create a user profile, which allows you to post and share messages, updates, photos or weblinks with your 'friends'. Members can send or accept 'friend' requests to other users; generally you can then see and comment on the activity of any friends, who can in turn 'like' or comment on your posts. You can also privately 'instant message' other members/groups. Advanced privacy tools allow for substantial control over public & friend access to your content.

Twitter: Registered users can read and send short posts or 'tweets' (limited to 140 characters). Tweets often include links to other tweets, Facebook or Instagram posts, or other weblinks. Tweets are public by default, but senders can restrict delivery to just their 'followers'; users can also restrict who can follow their account. Tweets from those you choose to follow appear in your feed. Posts can be grouped by topic or type by using hashtags, words or short phrases prefixed with a # sign.

Tumblr: A microblogging / social networking site where users can post multi-media content to a short-form blog. Users can also follow, comment on or re-blog the posts of other bloggers through their feed or 'dashboard'. Posts can be 'tagged' to help followers find specific content. All content is public but users can require a password to access their blog.

Instagram: Members can share photos with their followers, including comments about the photo. Followers can 'like' or comment on the photo posted, or share it on to their followers or through other social media platforms. Users can require prospective followers to obtain permission to view their posts.

Pinterest: This platform allows members to 'pin' and share links to interesting web content they have encountered, or to save the existing 'pins' of other members. Clicking on any 'pin' will take you to the original website for the pinned content. Members can organize their pins into categorized boards. Other members can choose to follow you but your permission is not necessary.

Loebach, J. & Madigan, R. (2015). *Collecting Social Media Data for Qualitative Research* (Research Shorts Series #2). Charlottetown, PE: Young Lives Research Lab, University of Prince Edward Island.

Collecting Social Media Data

Creation of a research team profile

In order to be able to view a participant's private social media activity, most platforms require you to be accepted by the participant as a 'friend' or a 'follower'. We suggest creating a neutral online profile for the research team, using the same profile name across all platforms, with which to either "friend" (Facebook), or "follow" (Twitter, Instagram, Pinterest, & Tumblr) each participant.

Connecting to participants online

During the initial interview, the team can ask each participant for permission to collect their social media data for any or all of the selected accounts they use regularly (note: the popularity/relevance of any platforms selected for inclusion should be reviewed regularly). Those who consent then provide their usernames, in writing, for each platform; the research team profile is then used to search each platform for the username provided, then to request permission (as necessary) to friend or follow participants. In some cases, it can still be difficult to locate a participant, or to confirm you have the correct account, if the information provided is incomplete or if there are several accounts with the same or similar username. Verification can sometimes be provided through comparison with a participant's other platform profiles (e.g. they may use the same profile photo for multiple platforms, or provide links to their other social media accounts). Alternatively, if a friend/follow request has already been accepted on another platform you might message them through the other medium to confirm

account information. To avoid these issues, we recommend having participants initiate or accept the friend/follow request directly during the initial interview. This ensures you connect to the correct participant, and provides another layer of consent.

If possible, we recommend that you capture all publicly available data from each participant prior to friending/following them; this provides general socio-demographic information, but also allows researchers to contrast information available to friends/followers versus that available to the public.

In some cases, participants may not respond to the friend/follow request initiated by the research team. This may mean they changed their mind about sharing their online activity, or are no longer interested in participating in the study. If a participant does not accept the friend/follow invitation after two requests, we recommend ceasing the requests but perform a collection of the public data from their social media profiles.

Ensuring participant privacy

To mitigate crossover between participants' social media accounts, the strictest privacy settings should be employed for the research account within each platform. Whenever possible, make the "stream" content available only to the research team so participants can not see or access each other's activity. This is especially important for Facebook, where the basic functions of the platform allow users to view one another's profiles, posts, and friend's posts.

'Capturing' Social Media Data

Social media data can be collected using a program extension by *QSR* called '*NCapture*'. Once installed, the *NCapture* extension appears as an icon on the URL toolbar of your web browser, and can be used to capture and code or name any web content. Depending on the platform, social media activity can be captured in a format similar to a 'screenshot' which captures content as it is formatted on the screen (saved by *NCapture* as a PDF) and/or as a spreadsheet-style dataset. Dataset files typically include metadata such as sociodemographic information (e.g. relationship status, religion, home town) and GPS coordinates for the location where the content was generated. Captured content can then be imported into *NVivo* and organized under 'Sources'. We recommend creating a folder for each participant, with subfolders labelled by the date of each collection.

The *NCapture* function is historical in nature; it will capture data back to the date on which the platform account was created. Unlike screenshots, *NCapture* will also collect what appears off screen; however, this is limited to content which has been 'loaded up' by the social media site itself. For example, Facebook allows users to scroll down through past posts in order to view content. To capture this earlier data with *NCapture*, one must scroll back through a participant's pages to the desired date to load all data back to this date. For frequent collections, simply scroll back to the date of the last collection.

NCapture will collect text, photos, video, and hotlinks (clickable URL links to web pages included in posts) in both the PDF and dataset formats. Hotlinks will remain active but any links to private content on a participant's accounts will only work while you are a friend/follower; once the relationship is terminated, only publicly available data can be viewed.

Note: Social media applications are constantly changing in style and format. There may be some data loss while NCapture is reprogrammed to accommodate changes; recommended procedures and timelines for data capture outlined here may also need to be adjusted.

When the collection period is over, researchers must actively 'unfriend' or 'unfollow' all participant accounts. Make sure participants are aware of this step from the onset to avoid confusion or insult. Remember too that the reciprocal nature of social media means collections will also include posts from participants' 'friends', whom did not explicitly give consent. All collected data must therefore be kept strictly secure and confidential; use of social media data for any academic or public venue should remove any textual or visual data that could identify participants' friends or contacts.

Recommended Guidelines for Data Capture

Which content is collected, and how often, should be dictated by research aims and whether or not research is longitudinal in nature. The frequency of the collection will inform the depth or richness of the data collected. Occasional collection of each social media feed can highlight the type and frequency of changes a participant makes to their profile, images, and friends/followers. Detailed notes should be kept throughout the collection period. However, as general rules of thumb we suggest:

Facebook: **PDFs**: first capture back to starting date of collection period; subsequently 2X per month; **Datasets**: 1X at halfway point, 1X at end of collection period.

Twitter: **PDFs**: *NCapture* collects full history of activity so complete data can be collected with only 1 collection per month. **Datasets:** 1X at halfway point, 1X at end of collection period.

Instagram: **PDFs**: As content changes frequently, aim to collect data 2X per month. Note that Instagram is a mobile app-based program so to access an account on a computer, type the participant's username into the URL address bar in a web browser (e.g. Instagram.com/username). **No dataset available**.

Tumblr: **PDFs**: Content changes frequently so aim to collect 2X per month. Note that users may create several pages within their blog; each page must be captured individually. **No dataset available.**

Pinterest: **PDFs**: Content changes frequently; collect 2X per month; capture screenshots of each 'board', and of all pins as a whole. A participant's 'likes' can also be collected but this creates a large dataset and can create privacy concerns (it will capture content from other people's pages). **No dataset available**.

We recommend that you collect a dataset from all platforms halfway through the collection period, in case the participant severs the relationship before the end of the study; if this occurs, one can still collect publicly available data, but detailed activity PDFs and datasets will no longer be available.

Note too that captured data for some platforms may extend beyond the approved collection period; PDFs and datasets may need to be truncated.