

## Human Subjects Considerations for Social Media Research

*Relevant for applications to PAR-22-164 and -165: Innovative Approaches to Studying Cancer Communication in the New Information Ecosystem (R01 & R21 Clinical Trial Optional)*

### Introduction

The massive trove of data available on social media platforms provides researchers with an unprecedented opportunity to gather and analyze data that would otherwise be very difficult, expensive, or time consuming to obtain. While social media data can help answer many important research questions pertinent to health, it is vital to ensure that these data are collected and used in an ethical manner. Although researchers are always expected to adhere to the general principles laid out by established ethical frameworks (e.g., the Belmont Report), these guidelines were written before the advent of social media, and don't address some of the unique features of these platforms. More recently, the updated Collaborative Institutional Training Initiative (CITI) Social & Behavioral Educational Modules Basic Course, which NIH-funded investigators are required to complete, includes content about social media and makes a clear distinction between the use of the Internet as an environment to study human behavior and the use of the Internet as a research tool.

There is currently a lack of consensus in the field regarding the proper way to collect and analyze social media data. This has led to great variability in the approaches that researchers take in designing studies, and that IRBs employ in evaluating social media research [1]. While determinations regarding any particular study are ultimately at the discretion of the researcher's individual IRB, for the purpose of NIH grant applications, it may be helpful to consider Department of Health and Human Services (HHS) guidance on what constitutes human subjects research and how that might apply in the context of social media data.

In this document, we highlight some key considerations for social media research, focusing exclusively on descriptive, observational studies that collect and analyze user-generated content from social networking sites such as Instagram, Twitter, and Facebook, as well as blogs, chatrooms, and forums [2], rather than studies that use these platforms for experimental and quasi-experimental studies, as the human subjects determinations in situations where there is intervention or interaction with participants involve a different set of considerations.

### Human Subjects Research

According to [Protection of Human Subjects, 45 CFR § 46 \(2018\)](#) (known as the revised "Common Rule"), a human subject is "*a living individual about whom an investigator conducting research: a) obtains information or biospecimens through intervention or interaction with the individual (and uses, studies, or analyzes the information or biospecimens); or b) obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens*" [emphasis ours] [2]. Most observational, descriptive social media studies do not employ methods that require direct intervention or interaction with participants (part "a" of the definition), but typically do involve the collection or analysis of information about living individuals (part "b" of the definition). In these situations, assessing whether the information collected is "identifiable" and "private" is key to determining whether the proposed study meets the criteria for human subjects research and would therefore be subject to informed consent procedures and the implementation of specific protections for participants. It is important to note that the data being collected must be **both** identifiable **and** private to meet the criteria for human subjects research (see Figure 1).

### *Identifiable Information*

Information is considered *identifiable* under the Common Rule if "*the identity of the subject is or may readily be ascertained by the investigator or associated with the information*" [45 CFR 46.102(f)]. Determining whether data are considered "identifiable" under this definition may be somewhat challenging because a) the phrase "readily ascertained" is subjective and not clearly defined in the Rule, and b) research that involves the aggregation of multiple data sets may lead to unanticipated opportunities for specific individuals to be identified [3]. Nonetheless, if an investigator obtains de-identified social media data (e.g., from a commercial company) and will have no access to identifiers, or if an investigator plans to use data from a message board where all users post anonymously, the research would not qualify as human subjects research because none of the data are identifiable.

### *Private Vs. Public Data*

*Private* information is defined in 45 CFR 46.102(f) as "*information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record)*" [3]. Based on this definition, if individuals intentionally post information on publicly accessible social media sites, such information would be considered **public** unless existing law or the privacy policies/terms of service of the entity hosting the information indicate otherwise [3]. For example, if an investigator proposes to analyze videos available on YouTube and will not interact with the individuals who posted the information, this would not qualify as human subjects research because although the posts may be identifiable, the data have been generated from public, not private, materials [4]. Essentially, if the content can be accessed by any Internet user without specific permission or authorization being required from the individual who posted the information or from the entity controlling access to the information, the post is considered "public" and therefore does not meet the "*identifiable private information*" criteria for human subjects research [3]. While this is a useful precept, it is also important that investigators take into account the expressed norms, intentions, and expectations of any given virtual space in determining whether the data should be regarded as "public" [3].

### *Completing the Human Subjects and Clinical Trials Information Form*

If an investigator, in consultation with their IRB, determines that their proposed observational social media study is not human subjects research (because there is no intervention or interaction, and because the data are public or non-identifiable) (see Figure 1) they do not need to fill out most of the Human Subjects and Clinical Trials Information form (Form G. 500). However, the investigator does need to include a justification to explain why the application does not constitute research involving human subjects. If a study does not meet the definition of human subjects research, the investigator should answer "No" to "Are human subjects involved?" on the *R&R Other Project Information Form* (Form G.220). This information will then transfer to the *Human Subjects and Clinical Trials Information Form* (Form G.500). Therefore, the answer to the first question on Form G.500 ("Are human subjects involved?") will also be "No." The investigator should then answer "Yes" to next the question on the form – "Does the proposed research involve human specimens and/or data?"—and provide an explanation for why the application does not constitute human subjects research (even though it does involve collecting data about humans). This justification should be attached to the form as a PDF file and may include: information on the individuals and entities that provide the social media data, a

description of the identifiers that will be associated with the data (e.g., usernames or Twitter handles), a description of who has access to the participants' identities (most likely the general public or anyone with Internet access), and information about the manner in which the privacy of research participants and confidentiality of data will be protected. Once this justification has been attached, the rest of the Human Subjects and Clinical Trials Information Form can be left blank unless otherwise directed by the FOA.

### Risk/Benefit Considerations

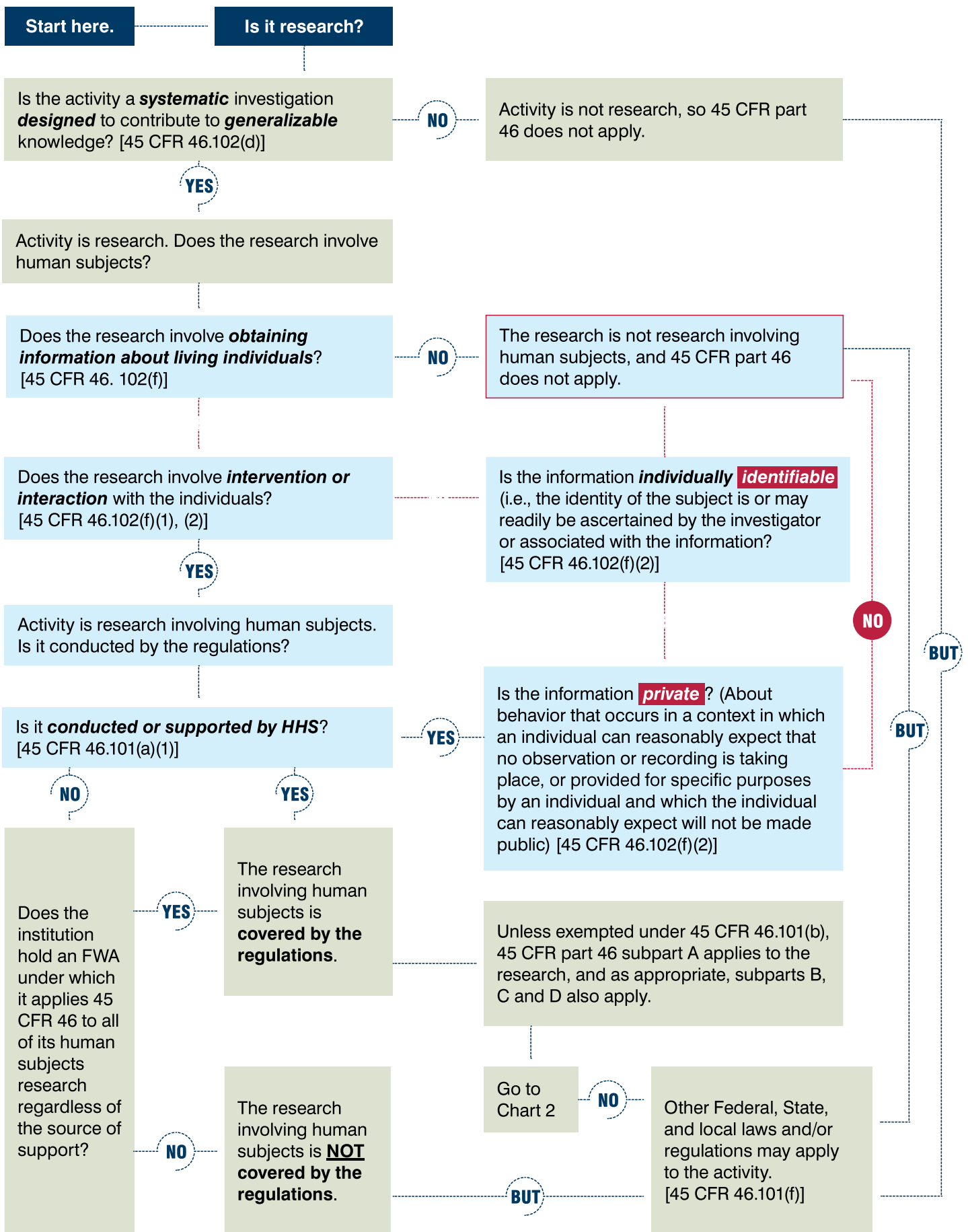
The potential benefits and harms of any proposed study must be assessed. This remains true for social media research, even though in many cases the potential harms are anticipated to be very low. [5]. Clearly, higher-risk studies (such as those that might expose a social media user to risk of embarrassment, reputational damage, or legal prosecution) require greater scrutiny than those that pose minimal risk. However, both the *probability* and the *magnitude* of harm should be considered when assessing risk - a possibility of harm that is of small likelihood may still be considered acceptable if the research team follows data security best-practices and takes steps to minimize potential vulnerabilities [3]. While the potential risks of social media research must be carefully considered and mitigated, it is also important to acknowledge that social media offer substantial opportunities for cutting-edge research, and these potential benefits must be taken into account as well.

In general, investigators conducting studies that do not meet the criteria for human subjects research (because the data they propose to use are either "public" or "non-identifiable") are encouraged to follow best practices in regard to data safety, and take precautions to mitigate potential risks (these steps may be outlined briefly in the data safety monitoring plan (DSMP) section of the NIH grant application). Additionally, although designating observational social media studies that involve public or non-identifiable data as "not human subjects research" is justifiable under the HHS Common Rule, investigators are strongly encouraged to consult with their individual IRBs to ensure that they are following best practices and are obtaining the most up-to-date guidance on these constantly evolving issues.

### References

1. Hunter, R. F., Gough, A., O'kane, N., McKeown, G., Fitzpatrick, A., Walker, T., . . . Kee, F. (2018). Ethical issues in social media research for public health. *American journal of public health, 108*(3), 343-348.
2. National Institutes of Health - Office of Extramural Research. (2019). Definition of Human Subjects Research. Retrieved from <https://grants.nih.gov/policy/humansubjects/research.htm>
3. Secretary's Advisory Committee on Human Research Protections. (2013). *Considerations and Recommendations Concerning Internet Research and Human Subjects Research Regulations, with Revisions*. US Department of Health and Human Services.
4. Moreno, M. A., Goniou, N., Moreno, P. S., & Diekema, D. (2013). Ethics of social media research: common concerns and practical considerations. *Cyberpsychology, behavior and social networking, 16*(9), 708-713. doi:10.1089/cyber.2012.0334
5. Benton, A., Coppersmith, G., & Dredze, M. (2017). *Ethical research protocols for social media health research*. Paper presented at the Proceedings of the First ACL Workshop on Ethics in Natural Language Processing.

**Figure 1. Is an Activity Research Involving Human Subjects?**



## G.220 R&R Other Project Information Form

**RESEARCH & RELATED Other Project Information** OMB Number: 4040-0001  
Expiration Date: 10/31/2019

1. Are Human Subjects Involved?  Yes  No

1.a. If **YES** to Human Subjects

Is the Project Exempt from Federal regulations?  Yes  No

If yes, check appropriate exemption number.  1  2  3  4  5  6  7  8

If no, is the IRB review Pending?  Yes  No

IRB Approval Date:

Human Subject Assurance Number:

**If conducting observational social media research (monitoring, surveillance, content analysis) with no enrollment, mark "no," per Figure 1 flow chart.**

**Source:** [https://grants.nih.gov/grants/how-to-apply-application-guide/forms-e/general/resources/images/newforms2017/newformpdfs2017/g220\\_otherprojectinformation\\_2017.pdf](https://grants.nih.gov/grants/how-to-apply-application-guide/forms-e/general/resources/images/newforms2017/newformpdfs2017/g220_otherprojectinformation_2017.pdf)

## G.500 PHS Human Subjects and Clinical Trials Information

### PHS Human Subjects and Clinical Trials Information

OMB Number: 0925-0001  
Expiration Date:03/31/2020

Please complete the human subjects sections of the Research & Related Other Project Information form prior to completing this form.

The following items are taken from the Research & Related Other Project Information form and displayed here for your reference. Any changes to these fields must be made on the Research & Related Other Project Information from any may impact the data items you are required to complete on this form.

Are Human Subjects Involved?  Yes  No

Is the Project Exempt from Federal regulations?  Yes  No

Exemption number:  1  2  3  4  5  6  7  8

The answer to the first question (“no”) will be carried over from the R&R, above.

### If No to Human Subjects

Does the proposed research involve human specimens and/or data?  Yes  No

If Yes, provide an explanation of why the application does not involve human subjects research.\*

Add Attachment

Delete Attachment

View Attachment

Skip the rest of the PHS Human Subjects and Clinical Trials information Form.

**\*The explanation for “not human subjects research” should include that the data are public, not private, and therefore do not meet the “identifiable private” criterion outlined in 45 CFR 46 (the Common Rule) to be considered human subjects research. Regardless of this designation, investigators should use this attachment to demonstrate ethical practice and due diligence toward protecting privacy, such as aggregating individual data, anonymizing when possible, safeguarding sensitive data, not publishing PII, paraphrasing instead of using direct quotes, and putting restrictions on data distribution and use.**

The answer to this question would be “yes,” because the research involves data from humans.

### If Yes to Human Subjects

Add a record for each proposed Human Subject Study by selecting ‘Add New Study’ or ‘Add New Delayed Onset Study’ as appropriate. Delayed onset studies are those for which there is no well-defined plan for human subject involvement at the time of submission, per agency policies on Delayed Onset Studies. For delayed onset studies, you will provide the study name and a justification for omission of human subjects study information.

### Other Requested Information

Add Attachment

Delete Attachment

View Attachment