

# Building a Framework for Social Media Data Reuse in Public Health Research

Lorinette S. Wirth, MPH

"Pay close attention to the content of social media access, production, and education...shift from narrowly focusing on the brightest signals of social media to highlight those that are obscured...make a case for the scientific need of social media data access in the name of public health and social welfare. We will need industry- and university-based researchers to unite."  
- Gray M.

## BACKGROUND

- Social media generates massive amounts of patient-reported data.
- Unique challenges arise when working with social media data, as is common with other sources of secondary data.
- Scientists must adopt innovative and multidisciplinary approaches to utilize this data for improved healthcare quality.
- Consensus as to where and how social media data utilization fits within public health research design is needed.
- Theoretical frameworks can serve as a roadmap for the reuse of patient-generated data from social media for conducting rigorous, evidence-based research.
  - leads to better formulated research questions and hypotheses, more appropriate selection of research methods, and conclusions
- Recent attention has been placed on online health communities (OHCs), a subset of social media where patients and their caregivers gather to learn about a shared illness, seek and offer support, and connect with others in similar situations (e.g., "Cancer Survivors Network").

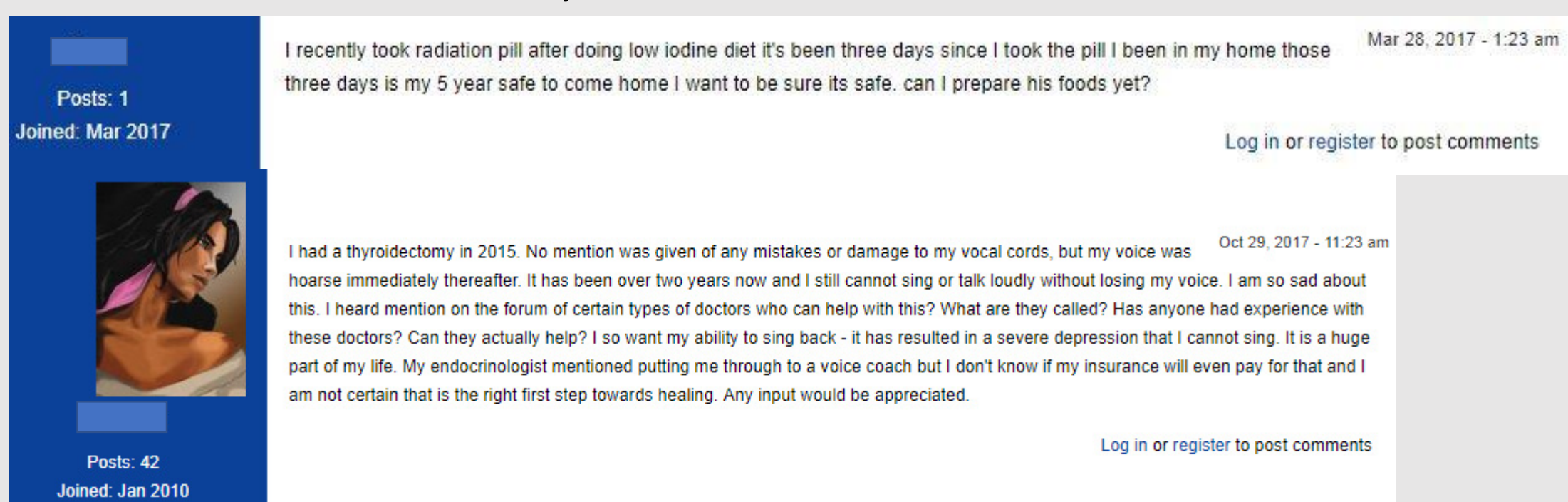


Figure 1. Examples of posts from a cancer OHC

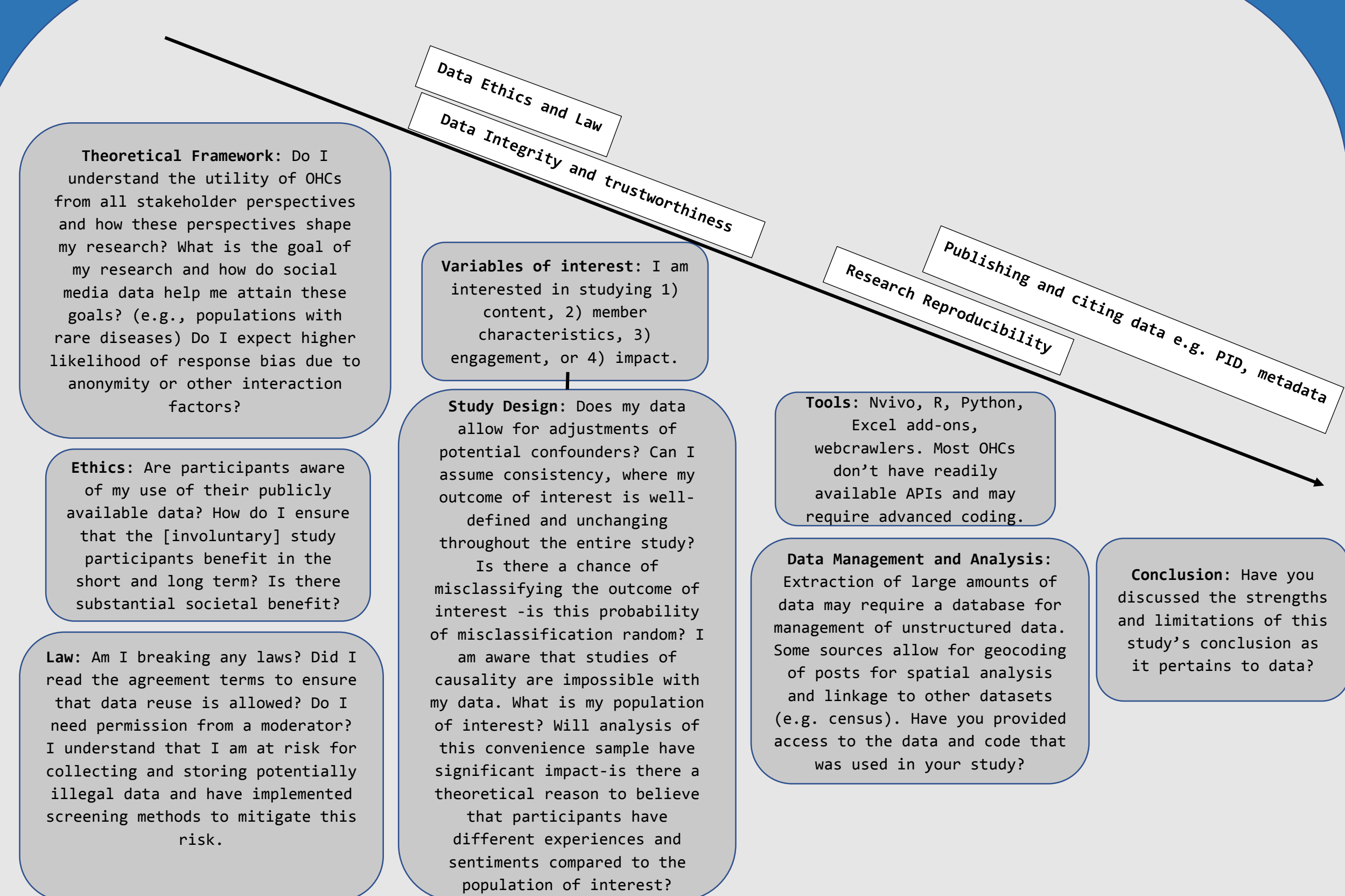
- **Objective: to develop a framework with recommendations for the reuse of patient-generated data from OHCs in public health research**
  - Insights gained are expected to be mostly generalizable to other social media sources and will serve as a foundation for future studies

## METHODS

- Queried PubMed database using search term: ("medicine"[MeSH Terms] OR "medicine"[All Fields]) OR "public health"[All Fields] AND "Social Media"[Mesh] AND hasabstract[text] plus snowball sampling method
- Inclusion criteria included articles that 1) had available peer-reviewed abstracts that were published after 2005, 2) reported analysis of social media data within (bio)medical or public health domains, 3) were written in English, and 4) were original research.
- We reviewed abstracts (n=300) and excluded studies that described disease surveillance, health interventions, health messaging strategies and evaluations, and use of OHCs for recruitment purposes.
- Recommendations and guidelines from the the Research Data Alliance (RDA) working and interest groups were incorporated into framework.

## RESULTS

Table 1. Description of a sample of reviewed articles				
	DGiles & Newbold	Song, Mei et al.	Thompson, Vázquez-Otero et al.	Lu, Yingjie et al.
Human Classification		✓		
Rhetorical analysis	✓			
NLP		✓		
Thematic/ Content Analysis		✓	✓	✓
Sampling				
Patient Consent/ Terms of service	✓			
Tools used		Python; Nvivo; MySQL		Webcrawler software (Offline Explorer 6.8)
Hypothesis driven/generating	Hypothesis generating	Hypothesis driven	Hypothesis generating	Hypothesis driven
Purpose	explore how diagnosis functions in the context of online mental health community forums	identify whether dentists discuss the oral-systemic connection and what aspects they discuss; to understand their perceptions of and attitudes toward the connection; and to determine what information they need to treat patients with systemic conditions (knowledge and attitudes)	explore the preconception health messages that are publicly available online, and critically examined these messages through a women's health perspective	determine the characteristics of different stakeholders involved in health care social media to facilitate an understanding of health care social media use.



### Things to consider:

1. Consider removing names (real or username), photographs as data is stored outside of forum.
2. Obtain permission from the moderator.
3. Patient consent not needed; perhaps, recommended.
4. Discuss most of the questions raised in this framework with publication of results.
5. Topics related to health information seeking behavior and needs, attitudes and knowledge seemed to be more readily accessible.
6. Report persistent identifiers based on PID Information Types (PID) WG outputs. Also, Data Citation WG developed recommendation to ensure that dynamic data is reproducible at the moment of computation and on the most up-to-date version of the available data.
7. Create standardized metadata with data publication.

Figure 2. A guiding framework for the reuse of data from OHCs for health research

## CONCLUSIONS

- Social media data reuse has the potential to guide public health research and policy, if used in a methodologically sound manner.
  - Learn about patient's experiences, what works and doesn't work to improve healthcare quality
  - Not meant to replace other data source, but compliment existing data and research
- While there is possibility for noisy or inaccurate data, the significance of the availability of fast, large-scale, low-cost, and unadulterated data collection via social media data rise is undeniable.
- Public mistrust of social media build due to events (e.g., US elections), education level and other disparities, failure by scientific community to report science openly.
- Information needs/ seeking, knowledge and attitudes were mostly studied; studies often lacked details as mentioned in Figure 2.
- Knowledge for analysis of social media data can generate hypotheses that inform future studies and medical practice. Some studies were also "hypothesis testing."
- The framework that is presented in this current project is the first of its kind, addressing the gaps bought by the scale and nature of this type of data. It should be used and expanded upon by researchers wanting to conduct rigorous studies using patient-generated data from social media.
- Future studies should expand upon this framework from an interdisciplinary perspective and design quality control methods such as scoring algorithms to assess validity of social media data and research based on the issues raised in this framework.

## ACKNOWLEDGEMENTS

This project was sponsored in part through grants from the Alfred P. Sloan Foundation #G-2014-13746 and the National Science Foundation NSF ACI#1349002" via the RDA US Data Share Fellows program. I would like to thank my RDA mentor, Dr. Leslie McIntosh for her guidance and support towards this project.

## REFERENCES

- Zhang, S. 2016. Computational Approach to Characterizing Online Health Communities.
- McFarland L. & Ployhart R. Social media: A contextual framework to guide research and practice. J Appl Psychol. 2015 Nov;100(6):1653-77. doi: 10.1037/a0039244. Epub 2015 Jun 8.
- Cesare N., Grant C., & Hawkins J. Demographics in Social Media Data for Public Health Research: Does it matter? Retrieved from <https://arxiv.org/ftp/arxiv/papers/1710/1710.11048.pdf>
- <http://journals.sagepub.com/doi/pdf/10.1177/2056305115578683>
- Giles & Newbold. Self- and other-diagnosis in user-led mental health online communities. Qual Health Res. 2011 Mar;21(3):419-28. doi: 10.1177/1049732310381388. Epub 2010 Aug 25.
- Gray, M. Putting Social Media in Its Place: A Curatorial Theory for Media's Noisy Social Worlds. DOI: 10.1177/2056305115578683
- Syred, Jonathan et al. "Would You Tell Everyone This? Facebook Conversations as Health Promotion Interventions." Ed. Gunther Eysenbach. Journal of Medical Internet Research 16.4 (2014): e108. PMC. Web. 18 Mar. 2018.
- Song, Mei et al. "Are Dentists Interested in the Oral-Systemic Disease Connection? A Qualitative Study of an Online Community of 450 Practitioners." BMC Oral Health 13 (2013): 65. PMC. Web. 18 Mar. 2018. doi: 10.1007/s10995-016-2213-8.
- Lu, Yingjie et al. "Understanding Health Care Social Media Use From Different Stakeholder Perspectives: A Content Analysis of an Online Health Community." Ed. Gunther Eysenbach. Journal of Medical Internet Research 19.4 (2017): e109. PMC. Web. 18 Mar. 2018.
- Walker, A. Using Social Media to Identify Patient-oriented Barriers to Sickle Cell Therapies. 2017. Foundation for Sickle Cell Disease Research 11th Annual Symposium