Research Data Services

Data Literacy Series, n.05/2026

Key Considerations in User-Generated Content Research

Publicly available user-generated content (UGC), such as social media posts, comments, and forum threads, is often assumed to be freely accessible and reusable, but that assumption warrants closer scrutiny. Although UGC-based research is frequently exempt from Institutional Review Board (IRB) review, important gray areas remain regarding data handling, reuse, and redistribution, which require careful ethical and legal considerations.

While UGC is often technically public, frequently indexed by search engines, and accessible without authentication, its status as openly available is more complex than it appears. Such content is created within contexts shaped by social norms, platform affordances, and user expectations that do not map neatly onto traditional notions of "public record".

What to Keep in Mind

- "Public" describes access, not consent: people who post publicly have not necessarily consented to all downstream uses of their data, even when the platform's user terms account for the possibility of research.
- Harm can emerge at the dataset level: even when individual records carry no apparent risk, their combination may pose one.
- Ethical review is not only for sensitive categories: the re-sharing of public UGC data warrants review whenever aggregation, analysis, or vulnerable groups are involved.
- Support openness with guardrails: cite your data sources and detail your collection methods. Where possible, re-share de-identified or aggregated versions of the data.
- Data protection laws restrictions: for example, under the GDPR (Article 3), re-sharing identifiable user-generated content without remediation requires a lawful basis, and the scope is determined by the user's location, not the researcher's.

Before You Proceed

Know the Law

Stay current with platforms' Terms of Service, as well as local and international data protection regulations, which vary by country and continue to evolve.

Assess for Vulnerability

Special care is required when content involves vulnerable populations or sensitive topics. Even when data is sourced from public platforms, its compilation can heighten the risk of unintended harm.

Avoid Compounded Risk

When combined or structured, data may enable inferences or expose sensitive content invisible in individual records. This risk must be assessed before any data is shared, linked, or published.

Adopt Protection Measures

Where possible, data should be de-identified, minimized, or access-controlled to prevent re-identification of specific individuals. Avoid distributing clean, structured datasets that retain PII to minimize misuse.

DON'T FORGET! Always consult with the IRB to determine whether your study requires review or oversight.

Have questions? Consult with us:

rds@library.ucsb.edu

UC **SANTA BARBARA**
Library

www.library.ucsb.edu