

This chart provides a useful starting point for internet researchers to consider ethics. Convergence of technologies and capacities continues to break down the strict boundaries between these categories. While not intended to provide answers, it promotes consideration of a range of issues and questions that may become relevant in the course of any internet related research.

Types of Data collected	Types of Venues/Contexts	Commonly asked questions about ethical practice
<p>Interactions, behaviors, transactions</p> <ul style="list-style-type: none"> • Hyperlinks • Comments or Recommendations • File or Information Sharing (file or snippet) • Forwarding /Replying • Interpersonal Interactions, conversations • Networks (e.g., maps visualizing communication flow or strength of relations between persons) 	<p>Direct communication (formal or informal interviews via real-time or asynchronous text, audio, or visual)</p> <p>Special Interest Forums (email- or web-based conversations and archives, e.g., threaded discussion forums, chatrooms)</p>	<p>How is protection of autonomy of participant/author achieved through informed consent or protection of vulnerable persons? How can researcher ensure that author/participant understands and agrees that content or interaction may be used for research purposes? Is the communication archived or easily searchable and retrievable? Is the data subject to open data laws or regulations? How long does the third party provider or ISP preserve the data and where? Could privacy be achieved through anonymization of email content and/or header information?</p> <p>How do terms of service (TOS) articulate privacy of content and/or how it is shared with 3rd parties? Regardless of TOS, what are community or individual norms and/or expectations for privacy? Does the author/subject consider personal network of connections sensitive information? Is the data easily searchable and retrievable? If the content of a subject’s communication were to become known beyond the confines of the venue being studied – would harm likely result? Is the conversation thread or forum perceived as public or private by the author(s)/subject(s)? How is profile, location, or other personally identifying information used or stored by researcher? Is the data easily searchable and retrievable? How is informed consent or protection of privacy achieved? How are vulnerable persons identified and protected? If non-active archives are used, how is vulnerability or harm defined and how are potential or actual subjects protected?</p>
<p>Production, presentation, performance</p> <ul style="list-style-type: none"> • Texts (e.g., authored texts, naturally occurring discourse, interview transcripts) • Images (presented or produced by user or captured by researcher) • Video (presented or produced by user or captured by researcher) • Audio (presented or produced by user or captured by researcher) • User motions and movements (in online training, virtual spaces, produced or presented by user or captured by researcher) 	<p>Social Network Sites (e.g., LinkedIn, google+, Facebook, Myspace, Flickr, FourSquare)</p> <p>Personal sites/blogs (e.g., homepages, blogs, youtube, and all forms of multimedia presentation)</p>	<p>How do the terms of service articulate privacy of content and/or how it is shared with 3rd parties? Does the author/participant consider personal network of connections sensitive information? How is profile or location information used or stored by researcher? Does author/participant understand and agree to interaction that may be used for research purposes? Does research purpose and design balance possible conflicts between participant and researcher perceptions of public/private and sensitive/nonsensitive? Does the dissemination of findings protect confidentiality? Is the data easily searchable and retrievable? If the content of a subject’s communication was ever linked to the person, would harm likely result?</p> <p>Could analysis, publication, redistribution, or dissemination of content harm the subject in any way? If the content of a subject’s communication were to become known beyond the confines of the venue being studied would harm likely result? Does the author/participant consider personal network of connections sensitive information? Does author/participant consider the presentation of information or venue to be private or public? Do the terms of service conflict with ethical principles? Is the author/subject a minor?</p>
<p>Locations and movements</p> <ul style="list-style-type: none"> • Physical locations (GPS) • Physical movements • Surfing behaviors 	<p>Avatar-based social spaces , virtual worlds, and online gaming spaces (e.g., Second Life, SIMS, MUDS/MOOS, MPORPG)</p>	<p>Should these virtual worlds be considered “public”? What constitutes “privacy” in such places? Should avatars be considered as persons and afforded the same protections as human subjects? Will the process of requesting consent itself cause harm? How and when should consent be sought? What requires consent? To what extent do users perceive their interactions and communication to be private in these spaces? How do Terms of Service specify researcher presence, anonymity of users, and privacy/confidentiality? To what extent and in what ways could research activities interfere with or compromise a user’s play or outcomes in the game? How should researchers juggle their own multiple roles? Could data be used to identify a user’s physical location and other sensitive demographic information?</p>
<p>Archived information</p> <ul style="list-style-type: none"> • Demographic information • Bookmark collections • Discussion archives • Data banks • Transaction logs • Clickstream data • Trace data 	<p>Commercial Web Services (e.g., Google, AOL, Yahoo, Bing, MSN, SurveyMonkey, Cloud Storage)</p> <p>Databanks/ Repositories</p>	<p>What are the participant/author’s expectations of privacy? Is the data easily searchable and retrievable? Is the data subject to open data laws or regulations? Does the service’s privacy policy contradict ethical principles? What measures safeguard data at the site of data collection? How long will the data be stored on the servers? Does this contradict the time frame indicated by the researcher or institutional policies? What happens to the data after the researcher completes work on the service? How are the data destroyed? How will cross-border data be handled if IP addresses are considered by one country to fall under privacy regulations?</p> <p>Where is the data stored? How long will the data exist in the repository? What consent is needed for subsequent data use? Does the remixing/mashing of data enable identification of individual or group identities or enable any additional risks to participants? In the case of shared data, what conditions were placed on data use by the original researcher, if any? Regardless of conditions, what ethical responsibilities may require consideration by later users? What mechanisms are in place to ensure appropriate data provenance and ownership? How will images/audio be effectively anonymized?</p>