**University of Southern California**

**Confidentiality and Anonymity**

Often there is confusion between confidentiality and anonymity. In our everyday use of the terms, we often use the words interchangeably. However, when it comes to considering how to use them when explaining research, the terms have a very precise definition. Using the correct word is especially important within the IRB and research process, and it is important for the researcher to know the difference. With the IRB process at University of Southern California, clearly understanding the difference between confidentiality and anonymity is important when filling out the IRB application, using the [informed consent form (ICF) template/information sheet](https://hrpp.usc.edu/irb/forms-and-templates/), recruitment of participants, and if you have a research assistant, transcriptionist, statistician, etc. helping with your process.

[(Please review this video to help understand the difference between an ICF and an information sheet).](https://www.youtube.com/watch?v=wJ2vokrFvnE)

**The difference between confidentiality and anonymity at a glance:**

**Confidentiality** is when you know who your participant is (like if you interview them, they are in in a focus group, or you need their name as part of any type of online data collection (sometimes in a questionnaire, sometimes in a survey), but you change their name to something else. You are keeping who they are confidential. This can also include any type of PII (Personal Identifiable Information) such as their zip code, an ID number of any kind, city, county, where they work etc). In most cases, qualitative research involves data collection that is confidential, because you know who you are interviewing and/or you know who is in participating in the focus group.

Please note that survey tools are often collecting confidential data.

Often surveys like using Qualtrics, are not truly anonymous, since you may collect ISP addresses, or you have a place for participants to record contact information so that they can claim compensation or an incentive.

\*\*\* Confidentiality can be disclosed if participants disclose during the study that they will harm themselves or someone else.\*\*\*\*

**Anonymity** is when there are no identifiers to link to data to an individual. In the world of Human Research Protection Programs, and professional organizations like [PRIM&R](https://primr.org/), nowadays (2023) anonymity is almost impossible to guarantee. There is evidence that data can be tracked to an individual’s home by their smartwatch/step tracker. It is possible to set-up surveys (like in Qualtrics and/or Survey Monkey) to not-collect ISP addresses, but often there are other identifiers. In some apps like Google forms, email addresses may be collected. Be sure if you make this choice, that the data collection is anonymous, that you truly insure that the platform you are using is set up not to collect identifiers.

**Please when filling out iStar section 26 you make sure you are making the appropriate choices.**

**What mechanisms can be implemented to promote confidentiality?**

What can researchers do to mitigate the risk that confidentiality is compromised?

* Signed Informed Consent forms should be protected. If collecting them electronically, be sure to keep them separate from the actual data collected from the participant. If collecting hard copies, place the informed consent letters in one envelope and not attached to interview transcripts, questionnaires, or surveys. For example, consent forms have the participant’s signature and if the researcher staples it to the completed survey, then a name is linked to the response.
* Findings are written up in a manner so that readers cannot identify participants’ identity. This is particularly sensitive if research is conducted in a small geographic region or in a setting where it is easy to discern the identity of the participants. **Deductive disclosure** can happen when a participant can be identified even if their name is not used. This can happen if personal identifiable information is collected that could link a person to data.
* Deductive Disclosure is something researchers should take very seriously. Protecting the rights of research participants is very difficult as photos, voice recordings, and even quotes can identify an individual. Often researchers think that if they de-identify the participants, or assign the participants a pseudonym or number, that they are no longer about to be identified. However, an individual that has unique speech patterns or uses specific language, may be identified by others.
* Findings are written up in an aggregate manner where no personal identifiers are attached.
* Store all research materials (i.e., interview transcripts, completed surveys, consent forms, etc.) in a safe, locked place, accessible only to the researcher – this may be electronic or paper.
* These data and research materials should only be accessible to the researcher and other approved members of the research team.
* Pseudonyms or participants code should be used in qualitative research.
* Data should be saved on encrypted drives.
* All data records should be protected for a minimum of 3 years as per the federal regulations.
* Non-disclosure and/or confidentiality agreements are signed to protect participants (e.g., transcriptionists, research assistants, statisticians etc.).
* Organizations, agencies, and institutions from where participants are recruited or where data are collected are not disclosed. Disclosure is when somebody could figure out the name of organization by the description of the organization even if the name wasn’t used. Fictitious organization, institution, or agency names are used. It is advisable to discuss with participating organizations, agencies, and institutions if they want the name of their organization written up in your research. If this is an option, you want to explore, it must be clear in the informed consent form for the research participants.