**Working with your IRB:**

 **Obtaining consent for open data sharing through consent forms and data use agreements**

By: Jeffrey A. Shero & Sara A. Hart

Florida State University

**Consent Form Language**

There are several types of language commonly included in informed consent or assent forms that though offer good coverage for protecting the privacy of individuals in the study, may be restrictive to the point of not allowing open data practices. One such type of language are overly general, blanket statements like below:

“Data will be heard or viewed only for research purposes by the investigator and his or her associates.”

“Results will be available only to school personnel and to the [specific University] researchers.”

“All information will be kept for at least seven years in a secure location and only project staff will have access to it.”

“All the information will be kept in secure locations and all digital files will be password protected so that only project researchers can access it.”

 These examples all intend to provide a sense of privacy by ensuring that only project investigators will have access to said data. Often times, it is not the intention of researchers to state that all information will be handled like this, but instead only the copies of data with identifying information. It is important to avoid the general blanket forms of coverage if it is possible that the already deidentified data may be shared with other researchers or stored in an open-access repository. Instead, try altering the language to specifically discuss what will be done with original records or identifiable information. For the first example above, this would like below:

“**Original (paper) records and identifiable (electronic)** data will be heard or viewed only for research purposes by the investigator and his or her associates.”

 The simple addition of “original records and identifiable” offers a significant degree of clarity about which data will be stored in secured locations. Making this distinction can also be extremely important if the consent form includes plans for destroying data information. This still does not however say anything about what will be done with the already deidentified data. Instead, include an additional line or sentence explaining what will be done with said deidentified data like below:

“Original records and identifiabledata will be heard or viewed only for research purposes by the investigator and his or her associates. **Data with all identifiers removed may be used for future projects that focus on any topic and may be unrelated to this study.**”

 The addition of this new sentence now makes a few things clear. First, another copy of the data with all identifiers removed will be made. Second, this copy of data might be made available to other researchers. Third, the studies that someone’s data may be used for might be in areas of research they never expected. It is important to be upfront and clear about all of this when making data publicly available. It also should be noted that although the language above seems to put the researchers in the clear to share the information for open-access use, it still may be unclear to the individuals in the study what they are agreeing to. It is important to remember informed consent forms exist first to protect the individuals in the study, not as a barrier for researchers to navigate. Adding additional clarifying language can help ensure peace of mind for study participants and ensure maximal protections. Altering the sentence above one last time with clarifying language we get the statement below:

“Original records and identifiabledata will be heard or viewed only for research purposes by the investigator and his or her associates. Data with all identifiers removed may be used for future projects that focus on any topic and may be unrelated to this study. **This new data may be made available to the general public via the Internet and an open database. This information will not have your name or other personally identifiable information included (i.e. it will be de-identified). Therefore, the data we share with the general public will be free of information that would link your responses to it.**”

 What results now is a clear statement to both researchers and participants with specific information about expectations for identifiable data and deidentified data.

Confusing language

 Beyond the issues presented above, there are several other forms of wording to avoid. First, is avoiding any contradictory information. This may seem like common sense, but regardless of the amount of specific language in a document about sharing deidentified participant data online or with other researchers, having a single sentence to the contrary saying that all data will be destroyed or kept in private locations can uproot or overturn all of that clarifying language. Second, vague or ambiguous statements such as, “will be protected to the extent of law”, can prove trouble for researchers. If you are going to make such statements, be sure to be aware of what said laws say about data sharing and what constitutes deidentified data. Further, when possible be specific about which law authorizing bodies you a referring to. This is particularly important for studies that span multiple cities, states, countries, or other localities that have differing laws. Whereas you may have expectations about data sharing based on your own locality’s laws, participants elsewhere may have expectations for the laws in where they live. Thus providing specific language about which law authorizing body you mean or ensuring you meet the expectation of every law that may be considered for your specific study is necessary. Finally, ensure that this information is consistent across consent forms. Whether it is having multiple waves of data collection, data from multiple parties (parent, student, teacher, etc.), or if data is combined from multiple collections/different projects, it is important to ensure concision of language across the documents. Having clear language in one document does not make up for unclear language in another. Additionally, clear language about sharing data in one consent form does not necessarily make it clear if prior or following waves of collection may also be shared.

Other important points to consider

 Beyond the language listed above, there are other pieces of information that can be helpful for navigating your consent forms for open data use. One such piece of information is how internal review boards view data versus records. By IRB standards, records refer to the actual written or original reports of student measures and responses. These can be booklets, electronic records, physical tests, or come in many other forms and are often identifiable. Data on the other hand, is what results when these records are transformed into the datasets that we use for analysis and where deidentification typically takes place. It is important to note the distinction between these two as saying what you will do to one does not infer what you will do to the others. Another point for consideration is saying what you need and not simply following what many other consent forms have. For example, saying that you will destroy all identifiable data is often not the best choice for projects. This limits the ability to extend studies into more longitudinal studies or to add additional measures at a later point in time. Another example is saying that data will be identified by the extent required by the law is another commonly included phrase in consent forms, but one that is often not necessary and can lead to unnecessary confusion. Not including this unnecessary language can be helpful, but also may require some additional phrasing to take its place. For example, when not destroying identifiable data you will likely need to maintain an active IRB protocol for that identified data. When not including the phrase “as required the extent of the law”, you may need other more specific language that will help to ensure participants are still comfortable participating in the study. One final point to consider is that if your active consent forms do any of this and may not meet the standard for open data use, there is still hope. Waivers of consent exist and oftentimes can overrule any of the unclear or restrictive language in your original consent forms. If this is the case for your study, reach out to your IRB and inquire about completing a waiver of consent that will allow your deidentified data to be openly shared.

**Data Use Agreement**

 For studies that are yet to be conducted or approved by your IRB, including in the IRB specific plans for making data open/available to the public or outside researchers by request can potentially eliminate the need for a data use agreement. However, for studies that are ongoing or completed and that do not include a section in their original IRB protocol for sharing data, a data use agreement will likely be necessary.

 The data use agreement allows researchers to share what is known as a “limited data set” or a deidentified data set with outside researchers or with the public. The data use agreement must be signed by the individuals downloading or receiving said dataset, and should be first approved by your IRB if they deem it necessary. This document will serve as a contract between you, the university/institution, and the individual or organization receiving the data. It should include the date that data access was/will be granted as well as any direct stipulations/limitations an organization may have. An example of a data use agreement can be seen below.

“This Data Use Agreement, effective as of, [Date agreement goes into effect], is entered into by and between [Recipient of the data] and [Data holder/institution/organization]. The purpose of this agreement is to provide the recipient with access to a deidentified data set for the [Project name] project conducted by [Principal investigator name] for use in research projects outside of the original study.”

After this section, stipulations on the data use agreement can be inserted. This can include forbidding the data recipient from sharing data with other researchers until they sign a similar agreement or other stipulations that may be of concern.

 In addition to making those individuals downloading a dataset sign a data use agreement, you will also be responsible for signing a data use agreement upon uploading data. This data use agreement will ensure that you understand that data will be made available online, and can be taken down at any time but that data already downloaded or combined with other datasets cannot retroactively be taken down or removed in these instances.

“I understand that I may withdraw my data at any time before it has been anonymized and combined with other data. I understand that the anonymized form of the data I have provided will be made available to other researchers through publications and by being deposited in our data repository. I further understand that once data has been downloaded by a recipient, the recipient cannot be forced to destroy said data or be unallowed to publish with said data unless otherwise specified in a data use agreement.”

**Deidentified Data Sharing vs Identified Data Sharing**

Up to this point, the type of data we have discussed sharing has been primarily data that is already deidentified. However, there may be instances in which you want to share potentially identifiable information. Some instances in which this may be the case are sharing identified data with colleagues interested in contacting the participants to collect further data, sharing information that can be used to link to other variables (e.g. sharing school/district information to combine with publicly available data, sharing geolocation information for geocoding, etc.) or sharing quasi-identifying information that could potentially lead to reidentification. Sharing of data in this way **is not** the purpose of the language in this document. If this is your goal, a different and more explicit set of language will be required.

In these circumstances, researchers interested in sharing identifiable information should make this explicitly clear to participants. Consider including language like below to make this clear.

 “Identifying information may be shared with outside researchers who are interested in contacting you for follow-up studies or the collection of additional information.”

This language however may deter participants interested in participating. For example, participants may be concerned about you sharing their information with for-profit organizations. To address this concern, consider including either information about the specific researchers with whom the information may be shared, or state explicitly that information will not be shared with for-profit organizations or researchers interested solely in financial gain. Additionally, allowing participants to opt-in or opt-out of this type of data sharing can provide a balance of sharing this information without deterring a large number of participants.

 “Identifying information may be shared with Dr. Scientist at Generic State University who may be interested in contacting you for follow-up studies or the collection of additional information. Information will not be shared with private for-profit organizations interested in using your information for commercial research or private gain. If you would not like your information shared in this way but would still like to participate in this study, please check the box below.”

Beyond deterring participants, sharing in this way may not be possible as a result of other information within these consent forms. Although language related to destroying information after a given period of time is generally fine for sharing deidentified data (so long as you are destroying identifying/original records and not the data itself), the same cannot be said for sharing the information we are talking about here. Rather, stating that you will destroy this information after 7 years means that this information will be destroyed and no longer available. If it is your intention, or expectation at all, that you may share this information with anyone after this point in time then information such as this should not be included as it will directly preclude you from doing this. Additionally, if information will be shared prior to the date in which this information is to be destroyed, this should be clear as you no longer will be able to guarantee that this information will be destroyed. If this is the case and you are unable to remove this language for any reason, make this explicitly clear with language such as below.

“All original records and identifying information will be destroyed after 7 years. However, given that some of this information may be shared with outside researchers we cannot guarantee that. this information will not exist elsewhere once it has left our hands.”