**WRITING A DATA MANAGEMENT AND SHARING PLAN (DMSP)**

This is intended to provide you guidance on how to address the required elements of a DMSP, using the example of a hypothetical study to illustrate. Please review the Data Management and Sharing website for examples of specific plans for different types of research.

Example study:

Surveys will be administered to 100,000 adult California residents over the course of 6 months. The surveys will be collecting age, county of residence, race, COVID vaccination status, self-reported history of any positive COVID test (with or without symptoms) and history of any documented COVID related hospitalizations. All data will be de-identified and stored in RedCap. Data will be analyzed to determine how race effects vaccination rates and how that correlates with COVID infection rates. The results of this study will be entered into the California Health and Human Services Open Data Portal.

When designing a data management and sharing plan the following needs to be considered:

1. **What types and amount of scientific data is expected to be generated in the project**.
   1. This section should include a *summary* of the types and estimated amount of scientific data to be generated during the life of the project. In this instance, it is clearly stated above. This section should NOT contain preliminary analysis, case report forms, manuscripts drafts, data to be collected in future research, or any physical objects such as biospecimens or lab notebooks.
   2. Describe what scientific data will be preserved and shared, and the rationale for this decision. Will all collected data be preserved? If so, for how long? Will the data be destroyed at the conclusion of the study? Is there any rationale for not sharing this data? If so, explain why. The data resulting from this study will be indefinitely preserved and shared in an open-access database.
   3. List any metadata and associated documentation associated with the collected data, such that the study can be validated and replicated by another researcher. Metadata is “data/information about data". In this case it would be the study protocol and the survey instrumentation.
2. **Related Tools, Software and/or Code:**

In this section list by name all specialized tools, instrumentation, software and/or code necessary to access, manipulate and/or analyze the scientific data. This study is using Redcap as a tool to house and de-identify data. Any statistical software used during data analysis should also be listed here.

1. **Standards**

Data standards vary by discipline. They are a set of common rules that allow data and information to consistently and uniformly be shared and processed. An example in this case would be in what format should the data be recorded. For example, race should be recorded in a manner that is consistent with that of the field.

1. **Data Preservation, Access, and Associated Timelines** 
   1. Provide the name of the repository(ies) where the scientific data and meta data will be stored and archived. Be sure that any repository meets the NIH criteria for a suitable database. In this case the data will be collected and stored in RedCap. The data and metadata will be preserved and accessed through the California HHS Open Data Portal.
   2. Describe how the data can be found and identified. This is usually through a persistent unique identifier or other standard indexing tool. A persistent identifier is a long-lasting reference to the digital data. An *identifier*is a label which gives a unique name to a data or metadata entity.
   3. Describe when and how long the scientific data be made available. The data must be made publicly available no later than the time of publication (if the data is to be published) or at the end of the study. If the data will eventually be destroyed, describe when. The data from this study will be permanently open sourced.
2. **How will the data be accessed, distributed or reused?**
   1. Are there any factors such as consenting, privacy or HIPAA related concerns that would affect subsequent access distribution or reuse of data? In this case the IRB has approved the protocol that would sufficiently de-identify the data. There are no HIPAA defined personal identifiers being reported.
   2. Does this data need special controls? Will it be available to anyone who requests it or will approval be necessary? If the latter is the case, who is responsible for making that decision and controlling the access of the data in its repository? The data in this study will be provided through open access, so there are no controls as to who can access it.
   3. Discuss all protections of privacy, rights and confidentiality of human subject participants here. Describe how the data is de-identified. List IRB approvals and any Certificates of Confidentiality required. In this study the data was de-identified through RedCap.
3. **Oversight of Data Management and Sharing:**

Who is responsible for assuring compliance with this plan? How often will oversight take place? For example, will annual IRB review assure that this plan is complied with? Will The PI be responsible for assuring the plan has been followed at annual review?

DSMPs are specific to each project and discipline. Please refer to the sample data plans provided on USC’s Data Management and Sharing website for additional examples of other types of research and the appropriate data management and sharing plans for those specific projects.