

## Research Data Retention and Sharing

The importance of data to the outcome of the study for which it is collected is well known. However, data collected during research has value well beyond the specific aims of the immediate study. For example, after the study's results have been published, other researchers may suggest alternate ways in which the data could be analyzed which might lead to different or additional conclusions. There may be a need for access to data for retrospective regulatory review or to ensure the validity of the data and conclusions presented in a publication. Years after the publication of the results of a study, other researchers might wish to carry out meta-analyses in which data from multiple similar studies are combined to yield better statistical power, enabling confirmation of previous conclusions or allowing new research questions to be addressed.

Given the potential of research data for re-use, the federal government and many research organizations, including Weill Cornell Medical College, have established policies for the retention and sharing of research data after the publication of a study's main findings. Federal policies in place at the National Institutes of Health (NIH), the funding organization for the Clinical and Translational Science Center, require that a Data Sharing Plan be included in any grant application requesting more than \$500,000 per year in direct funds. In recognition of the value of research data for future use, and in order to ensure compliance with NIH and institutional requirements, the CTSC in turn requires that all research benefiting from CTSC resources (in the form of either direct funding or use of CTSC cores or other resources) agree to abide by NIH and institutional data retention and sharing policies. This is enforced through a written agreement that must be signed by the PI of each project before it can receive CTSC funds or use CTSC resources for the conduct of research.

Discussion of the NIH policy can be found at:

[http://grants.nih.gov/grants/policy/data\\_sharing/data\\_sharing\\_guidance.htm](http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm)

Weill Cornell Medical Center's policy is described at:

[http://weill.cornell.edu/research/for\\_pol/forms/data\\_retention\\_policy\\_final.pdf](http://weill.cornell.edu/research/for_pol/forms/data_retention_policy_final.pdf)

To **facilitate** data sharing, investigators planning for data management using the CTSC's REDCap system are encouraged to plan their datasets with sharing in mind. REDCap's REDLOC library offers electronic implementations of many standardized questionnaires commonly used in research studies. The REDCap management staff can also advise investigators on accepted standards for individual data fields, such as commonly used demographic form elements (e.g., gender, marital status, race, ethnicity).

Maintaining clear and thorough records of the conditions under which data have been collected (i.e., data about your data, or "metadata") is also key to enabling meaningful re-use of research data in the future. Investigators are strongly encouraged to retain not only their data but, at minimum, a textual description of the study design, intervention(s), and any other relevant facts about the context in which individual data points were collected.

National and repositories have been established for the sharing of genomic data, including [The Gene Expression Omnibus \(GEO\)](#) and [dbGaP](#). For advice and assistance in submitting final research data to these public repositories, researchers can visit one of the CTSC's weekly Walk-In Clinics. [WCMC Bioinformatics and Data Management Walk-In Clinics](#) are held weekly in the computer room of the Wood Library at Weill Cornell Medical College. [Bioinformatics Walk-In Clinics](#) also occur weekly at Memorial Sloan-Kettering Cancer Center, a partner institution in the CTSC consortium. Investigators from any of the CTSC partner institutions (Hunter College, WCMC, MSKCC or HSS) are welcome to attend clinics at either institution. Investigators unable to attend a Walk-In Clinic can contact [Alex Lash, MD](#), Director of the Bioinformatics Core at Memorial Sloan-Kettering Cancer Center, for further information.