Research Data Management 101

What is Research Data Management?

Research Data Management (or RDM) is a term used to refer to the process of organizing, annotating, preserving, and sometimes sharing research data. For our purposes, research data generally means information collected during the course of research processes used for analysis. RDM as a concept applies throughout the research life cycle in various ways. So, it is not just something you do once, but rather a process you put in place and apply continuously.

Why should I do Research Data Management?

Research Data Management can provide benefits throughout the research life cycle and beyond. Here are a few:

- *Streamlined Analysis*: data that has been collected in an organized and consistent way is much easier to use in analysis.
- *Validation & Replication:* data that is well managed can be more easily validated and results can be more readily replicated. Processes become more transparent and more easily understood.
- *Compliance:* Funding organizations are increasingly requiring data management as part of the award terms and conditions. It is expected sometime in late 2020 or early 2021 the NIH will begin requiring data management plans for all awardees.
- *Reuse:* well managed data can be more easily understood and reused by the original collector for additional analyses in two, five, or even ten years time.
- *Collaboration*: well managed data is more readily combined with other data and is easier to describe to and understand by parties not involved in the original collection effort.

How do I start implementing Research Data Management practices?

Although RDM practices are most easily managed when included in the research process from the start, RDM can absolutely be applied at any point in the research life cycle. Here are some basic practices you can start with:

- *File Naming*: determine naming convention(s) for your files. Different file types can have different conventions. Describe the conventions with examples in a readme file that is easily accessible to your team, and then apply these conventions during file creation
- *File and Folder Organization:* consider the types of data and files you will collect and generate throughout your process, and build a meaningful folder structure to house those items. Describe this briefly in a readme that your team can access, so everyone is on the same page
- *Pipeline / Analysis documentation*: record information about your various pipelines or analyses used to collect and process your data. If these are changing

Research Data Management 101

over time, apply versioning and record in a basic manifest which files are produced from which versions.

Where can I learn more?

Arcus offers additional Research Data Management services and resources to anyone interested. Contact arcus-support@email.chop.edu.