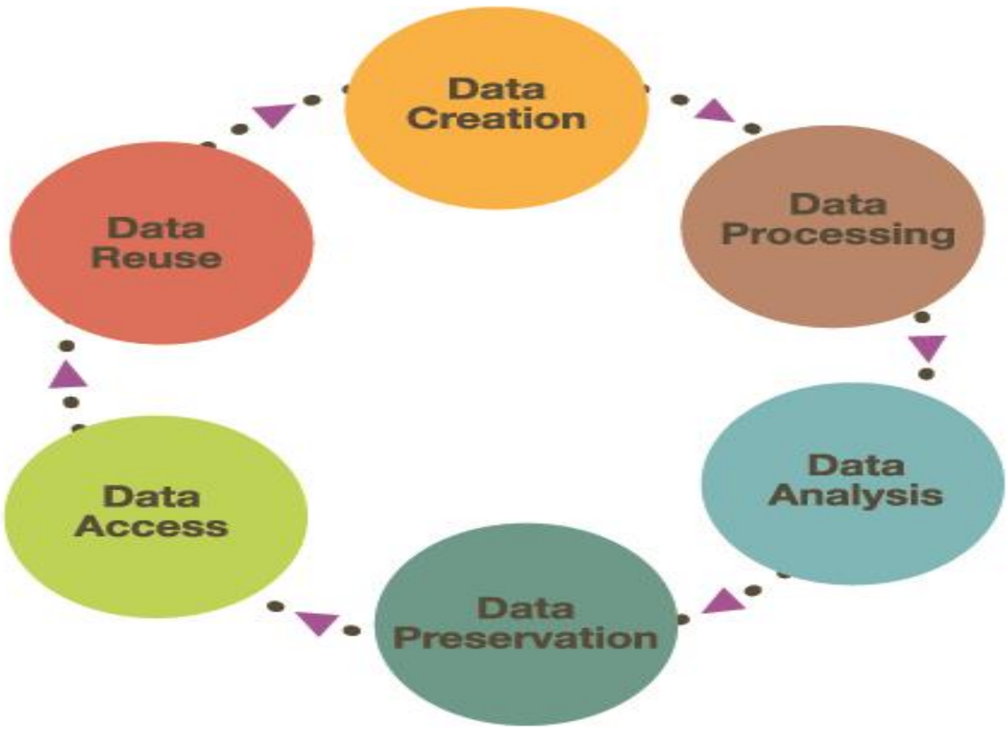
Topic:- Research Data Management:- Concept, Need and Features

Submitted to :- Prof. Rupak Chakravarty

Submitted by :- Raj Kaur Mlib,2(Sem)

(1)Introduction:- Research data management (or RDM) is a term that describes the organization, storage, preservation, and sharing of data collected and used in a research project. It involves the everyday management of research data during the lifetime of a research project. It also involves decisions about how data will be preserved and shared after the project is completed. Research data management involves the active organization and maintenance of data throughout the research process, and suitable archiving of the data at the project's completion. It is an on-going activity throughout the data lifecycle. research data management is a discipline concerned with making data generated in the course of research to be accessed as easily as possible by peers, contributors, and readers.

Research Data Management indicates the best practice when it comes to managing your data. It begins at the top with Data Creation and moves clockwise through the process. After data is created**,** it is then processed and analyzed. Data should then be preserved into archival formats and made accessible to the public. This will enable reuse by other researchers who will then create their own data and the cycle will begin again. Though this cycle assumes that each stage will take place in its entirety before moving onto the next stage, practice dictates that there may be several iterations of creation, processing, and analysis before it is ready to be preserved. Research data management describes a way to organize and store the data that a research project has accumulated in the most efficient way possible. It manages data gathered during the entire lifetime of the research project by coming up with consistent conventions. It is also responsible for the sharing, access, preservation, and secure disposal of data.



<u>(A) Research Data Life Cycle</u>:- **Plan:** Planning can include reviewing existing data sources addressing informed consent, considering costs, and preparing a plan.

Create: Researchers produce data (experiment, observation, measurement, simulation) and/or collect and organize third-party data and materials. Metadata and related materials are captured and created.

Process: Data is converted to digital format according to their quality assurance standards. Data is checked, validated, cleaned as needed. All these processes are documented, and the data is described using the appropriate discovery metadata standard.

Analyze: Data is interpreted and analyzed to produce research findings, publications, and intellectual outputs. Data sources are cited.

Preserve: Data is saved to formats that conform to curation best practices, user documents and discovery metadata are created, a digital identifier (i.e. DOI) is added and data is linked to any published products, consideration is given to security.

Share: Access rights are confirmed (ethics and intellectual property considerations). The data, along with user documentation and metadata, are made accessible, e.g. on a public domain server, or in a controlled repository.

Reuse: Potentially useful data, user documentation and metadata are located and obtained. Secondary analysis is conducted after any necessary data transformations are complete. Transformation are documented and data sources are cited.

There are four Types of Research Data:-

1. Raw Data 2. Curated Data 3. Published Data 4. Metadata

(B)Sources of Research Data :- **Observation.** This is data collected using observation of activity or behavior, usually through physical observation, surveys, or other sophisticated equipment like motion sensors, this data needs real-time documentation as it is impossible to “redo” or recapture again if not recorded.

- **Derivation.** This uses existing data to arrive at another piece of data, through extrapolation, interpolation, transformation, or some other method. An example is using available data from observation get to a conclusion, such as traffic volume. Unlike observational data, it can be redone, but it would cost a lot of time and resources to do so.

- **Experimentation.** This is what comes to most people's minds when faced with the term "research data." This is data that a researcher collects by changing variables to measure or look at differences in a hypothesis. This is particularly useful to find a cause-and-effect relationship and can be used via statistical projection to apply to a larger set. This is more reproducible than either observational or derived data, but still expensive.

- **Simulation.** Here, a test model is used to imitate a process or a system over time to find what would or could happen in several conditions. This model is often a computer-generated one, but researchers have simulated tests before using good old-fashioned pen-and-paper. Test models run the gamut of real-world systems, such as weather, geological activity, financial markets, neural pathways, chemical reactions, among others. What sets a simulation data apart is that the test model is usually more important than the test data. Depending on the type of model, simulation is a more affordable data source, though somewhat limited to the accuracy of the model—which is itself a conglomeration of data from other sources.

- **Reference.** Reference data, also called canonical data, is a type of secondary data source. This is a collection of smaller data sources, such as those above, or those already published and reviewed and are open for access or later research. Peer-reviewed journals, gene sequence databanks, or open-source code are some examples of reference data.

(C)Data Collection Methods :-

Whatever the source, researchers collect data using one of two methods: qualitative and quantitative. As you can surmise from the name, qualitative is descriptive, which is useful for things that can be observed but not measured. Quantitative, meanwhile, regards numbers.

Qualitative

Qualitative research is research that defines the associations of individuals and experience against a greater context, such as social realities or the world. It is more concerned with observing people and groups and how they live their lives in a particular setting. Therefore, qualitative research collects data that is more descriptive than empirical.

Quantitative

On the other hand, quantitative is a more objective type of method because it uses conventional standards of reliability and validity: numbers. While certainly not all data can be measured this way, quantitative data has the advantage of being categorized and/or ranked for a variety of purposes, such as graphs, charts, or tables. This property of quantitative data to be visualized shows the reader how to answer questions, not just display them.

(D)Challenges in RDM :-

Like anything else, managing research data has several challenges. The following are the biggest:

- **Improper storage of data.** This can lead to data being disposed of carelessly or become unusable. This is one of the biggest issues in data handling, which is directly caused by the research team's negligence. Depending on the terms of the agreement with the funder and/or sponsor, unusable data may actually be a violation. Plus, without proper data handling, inconsistencies will be overlooked.
- **Failure to document technical data.** Related to the first, this is a grave issue that stems from the team deviating from the proper standards of data documentation. This will make the findings of the research irreproducible, as any work that seeks to replicate the research will be riddled with inconsistencies.
- **The research institution gets no copy of the data.** Though rare, this is a big problem nonetheless if the original research team has left the institution and leaves them with no copy of the data. If later publications request access to the data—either to validate or recheck the findings of the report—it will put the institution in an awkward position. Maintaining data

elsewhere, such as personal servers, also presents problems, including legal ones.

(2) Needs of RDM :- There are several reasons RDM is important and much needed:- Data is a transient product and can be easily lost if not saved properly.

- Managing research data correctly saves time and money.
- Data that can be referenced, verified, and validated increases the accuracy and quality of the research. Sharing data can also often lead to developments and insights from its readers, even if they are outside the original research team.
- Managing research data helps spot errors, especially if data is accessible to your team.
- Funding agencies increasingly turn to data and reproducible results to approve research grants.
- Data, like journal articles and books, is a scholarly product.
- Data (especially digital data) is fragile and easily lost.
- There are growing research data requirements imposed by funders and publishers.
- Research data management saves time and resources in the long run.
- Good management helps to prevent errors and increases the quality of your analyses.
- Well-managed and accessible data allows others to validate and replicate findings.
- Research data management facilitates sharing of research data and, when shared, data can lead to valuable discoveries by others outside of the original research team.

(3) Features of RDM:-

- **Data Security**

  The most important benefit of RDM is that you can secure your data. By making an effective research data management plan, you minimize data loss and unauthorized access by adhering to data storage or organization standards. You also reduce the risk of losing the integrity of data either through accident .The most common site to store your research data is in your institution’s repository, like servers (for digital data). Your institution or organization may have advice on where to store your data.

- **Efficient Collaboration**

  The second most important benefit of RDM is collaboration, especially in an age where research is more complex, with more moving parts. But this is an advantage, as there is a positive correlation between the number of authors in a study compared to those with only one (Lamberts, 2013). Making data accessible for everyone in the group, even those not in the team but in the same discipline can open up massive opportunities to further your own research. Plus, good RDM routines also improve the efficiency of data access. An organized data directory structure, for example, can make contributing data or building upon the existing dataset much easier.

- **Organizing Data**

  Consistency and logic are the top two reasons researchers organize their data. It allows any member of the team to find and use them easily. You need not create a highly detailed flowchart for this, however, as it may simply entail thinking about a file naming convention and how to nest them in your directories for easy access. The ideal time to do this is before the project or the research begins. Naming conventions also preclude the possibility of overwriting files. File names may contain dates and other identifiers to help you track which files are yours and when they were modified. Metadata, however, is much more accurate for this task, which we'll also cover below.

- **Structure and Hierarchy**

  As mentioned, structuring your datasets in files and folders is an easy way to start your data management plan. Here are some ideas to get you started:

  Place files in the appropriate folder. Much like in real life, you would want to place files pertaining to a specific subject or topic in one folder. **Use hierarchy.** Use a few folders at the top for broader subjects, then more subfolders as needed for more specific topics.

Check for existing practices. If your team or institution already has a file structure and naming convention, see if you can adopt it so you don't have to start from scratch.

Stick to convention. In any case, stick to your file naming convention to prevent confusion, especially for newcomers to the team or the workspace.

Archive completed work. To streamline your work further, make sure superseded data is archived, not replaced completely. It may be useful to look at older iterations of research, such as to check for anything you might have missed. The important thing is to separate ongoing data from everything else.

Maintain a backup. Your data should be backed up, whether they are primarily saved on your local hard drive, on your intranet, or on the cloud. Your backups should have a backup, which means cloud storage, which syncs automatically with files on your local machine, is a great option.

- **Metadata**

  Metadata means data about data. This is information that tells you about the data contained in a file, which is helpful to find the exact file you are looking for (and for others too). At present, not only does metadata define data but it is also useful in bridging connections among tools and software, Metadata contains information that is necessary to find, interpret, and use your file, folder, or data. Like your file naming and folder structure conventions, deciding on metadata should be done at the start of the project.

- **Data Sharing and Preservation**

  Data will outlive the project, so you should plan for ways to share and preserve your data for posterity. Data preservation is part of the research data lifecycle. Though there are slightly varying models of data lifecycles (Ball, 2012), the research data lifecycle involves the movement of data from creation to preservation and reuse. Digital data has an advantage in the sense that it can be maintained for far longer than other types. However, the main drawback to this is that as technology progresses, the tools meant to access this data may change.

- **Backup**

  Data loss is the enemy of nearly every researcher—or near everyone who has stored files in any kind of storage medium. This is why it is crucial to have backups of your data, and to even backup your backup if necessary. Some institutions often use automatic backups to periodically save research work or any materials stored in their repositories. Ask your computer or network administrator for details of this automatic procedure, especially how often it happens, where it is stored, and how long the backups are kept. In any case, no matter how exemplary your institution's backup process is, it is still prudent to back your data up on your own.Cloud storage provides a relatively affordable but highly reliable means of backing up data. In addition, they offer competitive cost-to-space ratios. No matter the cloud provider, though, cloud storage syncs in real-time, so your remote backup data is updated as soon as yours do.

- **Manage Your Research for the Future**

  It can be said that good data management is not the destination, but the journey; it is how researchers lead to discovery and innovation (Wilkinson, 2016). Data, freely shared, can lead to further insights long after the original project is done and the research team has moved on. This is why it is important to have a logical data management system to index and store your research data, not only for your own use but for those who will come after. Citation is an essential part of the research environment, which brings your findings to the experts who can build on your work. Using initiatives like the DOI and new technologies, such as cloud storage can bring your research to more minds than ever before. To do that, however, managing your data just as your predecessors did is still a good idea. Following conventions, practicing logical data structure, and citing wisely is the framework upon which the future of science is built.

- **Sharing and Licensing**

  Researchers can share data using a variety of means. At its simplest, you can store them in a USB flash drive, which can be borrowed by colleagues. Otherwise, you can use FTP upload on a server, such as to your institution’s repository. Another way includes cloud sharing, which is explained below.As for licensing, investigators can simply make a request form that anyone who wants to use their data can fill out. Otherwise, if internet publication is preferred, Creative Commons licenses are ideal for research work.

- **Data Management Planning**

Knowing why you should manage your research data is all well and good, but the question remains: how should you do it? The answer is that you start with a data management plan, or a DMP, which will cover how your files and datasets are stored, organized, and arranged in a database. There are several database formats, which you can use for huge volumes of data.

Conclusion:- Research Data Management is a broad concept that includes processes undertaken to create organized, documented, accessible, and reusable quality research data. The processes involved in RDM are more complex than simply backing up data on a thumb drive and ensuring that sensitive data is kept secure. Managing data includes using file naming conventions, organizing files, creating metadata, controlling access to data, backing up data, citing data, and more. While many researchers are interested in data management, there are some who may not see a need for it at all. RDM is complex and imperfect because of the complexity of research itself. Supporting research throughout the data lifecycle by consulting with researchers and promoting best practices can be challenging, but will improve data quality, reproducibility, and shareability.

Reference :- https://www.guide2research.com/research/research-data-management

https://pitt.libguides.com/managedata