KOÇ ÜNİVERSİTESİ
SUNA KIRAÇ LIBRARY

2020

OPEN
ACCESS
Koç University

KOÇ UNIVERSITY
RESEARCH DATA MANAGEMENT
GUIDELINES

Çiğdem Yıldırım, Research Liaison

Koç University Suna Kıraç Library

7/18/2020

Contents

KOÇ UNIVERSITY RESEARCH DATA MANAGEMENT: GUIDELINES

Organizational policy on data management

Koç University shall require researchers to place into an archive that complies with international standards, the research data necessary to verify the results presented in scientific publications. Koç University requires that data and services be treated in an open fashion and in compliance with FAIR principles. The data should also be traceable and available whenever possible.

Koç University abides to the principle of Research Ecosystem: *"as open as possible, as closed as required."* If the data is not open for legal, confidentiality, or other related reasons (for example if it is sensitive data or personal data), this should be explained clearly. The metadata that makes the data findable shall be provided in all cases.

Koç University shall encourage the adoption of Open Science Cloud requirements for monitoring Open Science resources. Koç University shall provide an appropriate Data Management Plan for each research activity in which researchers are involved. Koç University researchers shall define their post-project usage rights by identifying appropriate licenses.

The minimum archival storage duration for Koç University research data is 10 years after the assignment of a permanent identifier in *Koç University Research Data Repository (KU-RDR)*. All legal and ethical perspectives of such actions need to be considered if such records are to be deleted or destroyed after the expiration of the necessary archival period or for legal and ethical reasons.

1. CREATING YOUR DATA

There are many decisions to make about managing your data before you even start creating/collecting it. These include data planning, choosing file formats, and addressing issues related to intellectual property rights, data protection and ethics. Many funders now require data management plans to be submitted with grant applications.

a. Making a Data management Plan

Therefore, your best starting point is to create a data management plan. A proper data management could be a useful resource to you as a guidance map to all your research datasets and can save you a lot of time during data collection, and also when consolidating your data towards the end of the project.

For the creation of data management plans we recommend using DMPonline templates. This software allows a stepwise construction of data management plans specifically tailored for your project and the requirements of various international research founders such as European Research Commission, Horizon 2020 and Welcome Trust. Using DMPonline is free of charge. Created plans can be exported in various file formats, e.g. for integrating them in project proposals.

Koç University Library has support staff who can help you create a data management plan for your project. Please get in touch openaccess@ku.edu.tr early, so there is enough time for the staff to

help. Koç University Data Management Plan template can be found at <u>Open Access & Scholarly Communication</u> webpage of the Library.

b. Choosing File Formats

In planning a research project, it is important that you consider which file formats you will use to store your data. In some cases, this will be dictated by the software you are using or the conventions of your discipline. In other cases you may have to make a choice between several options.

Best formats for preservation:

If you are not aware of any disciplinary standards these are some good file formats for the preservation of the most common data types:

Textual data: XML, TXT, HTML, PDF/A (Archival PDF)
Tabular data (including spreadsheets): CSV
Databases: XML, CSV
Images: TIFF, PNG, JPEG
Audio: FLAC, WAV, MP3, AIFF, WAVE
MS Access (.mdb/.accdb)
MS Excel (.xls/.xlsx),
MS Access (.mdb/.accdb),
dBase (.dbf)
OpenDocument Spreadsheet (.ods)
ASCII (.txt)
MS Word(.doc/.docx)
LaTeX (.tex)
JCAMP with JSpecView, ChemDoodle
Containers: TAR, GZIP, ZIP

c. Intellectual Property Rights

Sharing research data by depositing it in connection with a published article or otherwise making data publicly available sometimes raises intellectual property questions in the minds of depositing researchers, their employers, their funders, and other researchers who seek to reuse research data.

Intellectual property rights (IPR) management is an important part of any data management program. A builder of a database or other data resource will have an interest in who owns that resource and how others may use it. Someone who may populate that resource with data provided in part by others will want to make sure that all legal, ethical, and professional obligations that one may have to the provider of the data are met.

Since the benefits of data sharing are so well known and documented, a researcher may wish to share their database and/or content with others. Others can only fully utilize external data if they know the terms of use (if any) for that data.

- Data Ownership

As a researcher, you should clarify ownership of and rights relating to research data before a project starts. Ownership and rights will determine how the data can be managed into the future, so these should be documented early in a project through data management planning.

As part of their obligation to society in general, Koç University and its members have a responsibility to communicate the discoveries which are performed by scientific works and research to the public. Some of these discoveries may have commercial value that should be exploited to the mutual advantage of those concerned. If the use of these discoveries can be limited or controlled by physical or legal means, the discoveries may also be referred as "intellectual property (IP)."

Therefore, intellectual property, as used in this policy, includes not only technology such as inventions, discoveries, creations, or authored works which may be protected legally, such as patents and copyrights, but also the physical or tangible embodiment of the technology, such as specialized hardware, equipment or computer software.

The objectives of this policy are to:

(i)   encourage any member of the University who may have created or discovered intellectual property to disseminate that discovery to the public in a manner that benefits both the member and the University;

(ii)   (ii) recognize and uphold the principles of academic integrity in the possible commercialization of intellectual property;

(iii)   (iii) develop a body of knowledge and expertise within the University in order to permit the continued successful commercialization of intellectual property in the future;

(iv)   (iv) outline clearly the ownership rights in any newly created or discovered intellectual property as between the University and its members and the obligation for related costs and the division of related revenues;

(v)   (v) describe and define the rights and obligations of the University and its members in protecting and exploiting any newly created or discovered intellectual property; (vi) satisfy the requirements imposed by sponsors of the research;

[KOÇ UNIVERSITY POLICIES ON INTELLECTUAL PROPERY RIGHTS AND TECHNOLOGY TRANSFER](#)
governs all persons at the University, including full- and part-time faculty, visiting faculty, full-and part-time agents and employees (staff), full- and part-time students (both undergraduate and graduate), and fellows (pre- and post-doctoral), whether or not they receive all or any part of their salary or other compensation from the University.

- **University Ownership Rights**

1. Significant use of University resources: any substantial use of University laboratories, equipment, funds, personnel, or facilities, except those resources provided generally to all faculty and staff, such as the use of libraries and offices. Questions of whether someone has made a significant use of University resources will be resolved by the President (or with his or her consent Vice President for Research and Development) after reviewing RPDTTD's recommendations.
2. Institutional works: works created under the auspices of the University by employees or contractors as University. All "works for hire" as that term is defined under copyright law are

included. (Computer programs written by hired programmers are a good example.) Works which the University commissions non-employees to prepare, such as University publications, architectural designs, engineering studies, and consultants' reports, are also within the meaning of this term. In order to fully protect its interests, the University's policy is to obtain an agreement in writing whenever the University has an interest in owning a commissioned work.

d. Data Protection and Ethics

If you are working with people for your research project then you have a duty to ensure that any data you gather and subsequently use is handled correctly. Ethical guidelines are issued by funding organisations and also produced by Koç University.

In addition, laws such as the Turkish Personal Data Protection Law no.6698 which governs the processing of personal data, must be adhered to. The University provides guidance on this law and details the University's measures to ensure the regulations are met.

Internationally, Human Research Ethics Committees (HRECs) has published the document called "Data Sharing Considerations for Human Research Ethics Committees (HRECs)" which provides an overview for members of (HRECs) can consider when assessing applications which propose to share data.

2. ORGANIZING YOUR DATA FILES

Once you create, gather, or start manipulating data and files, they can quickly become disorganised. To save time and prevent errors later on, you and your colleagues should decide how you will name and structure files and folders. Including documentation will allow you to add context to your data so that you and others can understand it in the short, medium, and long-term.

a. Naming and organizing your files

Decide on a file naming convention at the start of your project. Useful file names are:

- consistent
- meaningful to you and your colleagues
- allow you to find the file easily.

Collect your data files under different names in separate files (e.g. Raw Data, Analysis Datasets, Final Dataset for Publication)

Example data filing and versioning:

Year_Month_Day_[File Name]_Version

2018_09_12_Weight_v1
2018_09_13_Weight_v2

File your data as soon as you create it, it can be hard to remember what you did at the end of the project.

Useful Renaming softwares for your files:

- Advanced Renamer (https://www.advancedrenamer.com/)
- [Bulk Rename Utility] (http://www.bulkrenameutility.co.uk/Main_Intro.php) (Windows; free)
- [Renamer](https://renamer.com/) (Mac; free trial)
- [PSRenamer](http://www.powersurgepub.com/products/psrenamer/index.html) (Linux, Mac, Windows; free)

Best Practices for File Versioning:

Versioning refers to savings new copies of your file when changes are made. This can be done manually or some software programs automatically do file versioning. Using file versioning can assist with keeping track of your data files as changes are made.

Simple file versioning:

- Include a version number at the end of the file name such as v1, v2, or v1.2, etc. Update this version number each time the file is saved.
- Include information about the status of the file, such as "draft" or "final. For the final version, substitute the word FINAL for the version number. Only save as FINAL when no more edits will be made.
- Include information about changes that were made, such as "original" or "cropped".

Note: When collaborating on a file, versioning control file naming conventions work best when all collaborators have agreed on the format early on.

Simple software solutions:

Google drive:

You need an Google account to use OHSU Box. Note- you do not need a gmail account, a google account can be created using any email address, including your OHSU email address.
New versions are saved for Google Drive's text documents, spreadsheets and presentations.
Version information includes who was editing the file and the date and time the new version was created. Google drive allows you to revert back to previous versions.

OHSU Box:
Use your OHSU credentials to access OHSU Box
OHSU Box tracks file versions for all files saved in Box.
The comments feature lets you indicate changes that have been made between versions.
Documents can be shared with others, and Box will track who uploaded or updated each file and when.

Additional information about OHSU Box:
OHSU Box has protections in place for OHSU confidential and restricted data or protected health information.

Storage size for OHSU Box is currently limited to 10 GB, with no requests for additional capacity accepted.

Additional software solutions:

Use version control software such as SVN, Github

b. Documentation and Metadata

To ensure that you understand your own data and that others may find, use and properly cite your data, it helps to add documentation and metadata (data about data) to the documents and datasets you create.

It is a good practice to begin to document your data at the very beginning of your research project and continue to add information as the project progresses. Include procedures for documentation in your data planning.

There are a number of ways you can add documentation to your data:

Embedded documentation

Information about a file or dataset can be included within the data or document itself. For digital datasets, this means that the documentation can sit in separate files.

Examples of embedded documentation include:

- code, field and label descriptions
- descriptive headers or summaries
- recording information in the Document Properties function of a file (Microsoft)
- README files

Supporting documentation:

This is information in **separate files that accompanies data** in order to provide context, explanation, or instructions on confidentiality and data use or reuse. Examples of supporting documentation include:

- Working papers or laboratory books
- Questionnaires or interview guides
- Final project reports and publications
- Catalogue metadata

Supporting documentation should be structured, so that it can be used to identify and locate the data via a web browser or web based catalogue. Catalogue metadata will be structured according to an international standard and associated with the data by KU Research Data Repository when materials are deposited. Examples of **catalogue data** are:

- Title
- Description
- Creator
- Funder
- Keywords
- Affiliation

3. DATA COLLECTION, STORING & SHARING

a. Data Collection

Koç University Research Data Repository (KU-RDR) has solutions for archiving, data sharing, if you wish the open access publication of your data.

All researchers affiliated with the university can use this service for archiving and publishing their data without any charge. KU Research Data Repository (KU-RDR) offers the following services:

- Persistent and citable addressing of data via DOI's
- Individual authorization and license models for data access
- Linking between research data and related publications
- Archiving of various file formats
- Tracking of version numbers of the data files
- Generating Data Cite as references
- Flexible and configurable metadata information
- Indexing of all data sets and respective data citation in the national and international databases such as HARMAN-TÜBİTAK, OpenAIRE and re3data.org.

Please contact openaccess@ku.edu.tr and give us some initial information about your data, especially about file formats and file sizes. Consider the future accessibility of your research data (open or partially / temporarily restricted access/ closed) and appropriate licenses - we can help you finding an optimal solution for your needs.

Research data files will be collected through the Library's Libwizard Data File Collection System by contacting the Data Stewards for the data publication gets a prefilled author contract he/she has to sign and we release the data. Essential file templates such as README files, Metadata files and Author Contract will be provided by the Library in advance.

Researhers who want to upload their data in the repository, are required to fill in these forms below related to their research data:

- Data management plans
- README files-related information about the datasets
- Metadata files-related metadata of the data
- Author Contract- to be signed by the principal author of the project.

Librarians will check your data and, if necessary, they carry out format conversions with regard to the long-term preservation of the data. We upload the data into the repository with related metadata information and prepare the publication.

b. Data storing and backup

You've invested a lot of time and effort in creating your data, so keep it safe. Learn how to select what to keep and how to store it carefully. Discover why and how to back it up to make sure it is not lost. Find out how to preserve your data and back-ups, and consider how you can get the most from your data, perhaps through re-use and sharing.

Storage:

Choosing the right way to store your data can help you work more flexibly, easily and quickly. Thoughtful storage solutions can also simplify version control and collaboration with others. No matter which solution you use, the two golden rules of storage apply.

1. Where possible, only store what you need to keep.
2. Store crucial data in more than one secure location.

Portable storage media such as memory sticks (USB sticks) and external hard drives are more risky and vulnerable to loss and damage. Computing officers will not back them up or support them centrally. It is important not to rely on them as your only copy of important data.

They are very convenient though, and useful for:

- temporary copies/moving files e.g. taking a presentation to a conference
- secondary or back-up copies
- files only one person at a time needs access to
- data you can afford to lose

Backup:

Nearly everyone who has experienced serious data loss did not think it would happen to them but it does happen periodically. The results can be catastrophic for your research project, or for you personally.

KU Research Data Repository (KU-RDR) which is a cloud-based storage provides a convenient way to store, back-up and retrieve data. There are many provides of cloud storage out there and you should check their terms of use before using them for your research data. The University provides three different cloud options – OneDrive,Google Drive and Sharepoint- which should be accessed with your University email address. If you are handling personal or sensitive data then you should check the cloud option is compliant with any data protection rules the data is bound by.

c. Sharing (Data Licensing):

One relatively simple way to make it easier for others to re-use tools, data, or other content that you produce is to add a Creative Commons license.

Creative Commons (CC) (http://www.creativecommons.org/) has a library of standardized licenses, and some of them apply to data. Attribution 4.0 International (CC BY 4.0), for example, is the equivalent of a Creative Commons Attribution license (CC BY) which is going to be applied to the data

in the ***Koç University Research Data Repository (KU-RDR)*** if the data does not have any limitations of personal or commercial privacy.

KU Research Data Repository accepts the [Attribution 4.0 International license](#) (CC.BY 4.0) which is a common Creative Commons license for research data– that

- you are free to share, copy and redistribute the material in any medium or format
- to adapt, remix, transform, and build upon the material for any purpose, even commercially.
- This license is acceptable for free cultural works.
- The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices: You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

There are other types of Creative Commons licenses, also allowing commerical use, and license which do not require re-user to attribute the creator. Creative Commons licenses are often used for materials released online, but you can also include these in printed materials if your publisher does not own the rights. For additional information about Creative Commons license options, visit [Creative Commons](#) website.

To license something with a Creative Commons license, you don't need to file any paperwork - just publish (in print or on the web) your materials along with a notification that you are using a particular license.

IMPORTANT NOTE: Creative Commons licenses are **'irrevocable'** so do not add a Creative Commons license unless you are sure that:

1. you have the right to publish this information
2. you will not want to revoke it later on for any reason.

d. Sensitive Data

Sensitive data are data that can be used to identify an individual, species, object, or location that introduces a risk of discrimination, harm, or unwanted attention. Major, familiar categories of sensitive data are:

- personal data / confidential data
- health and medical data
- ecological data that may place vulnerable species at risk

If your research involves human subjects, you should consider the [Turkish Personal Data Protection Law no. 6698](#) or if it has trade secrets you should use standards ISO 27001. Koç University has a law department and ethics committee to answer your questions about sharing sensitive data in your research projects.

KU Research Data Repository 's motto about open data is "as open as possible, as closed as required." If the data is not open for legal, confidentiality, or other related reasons (for example if it is sensitive data or personal data), this should be explained clearly. The metadata that makes the data findable shall be provided in all cases.

Publishing your data, or just a description of your data, means that others can discover it, reuse it and cite it. But, If necessary to limit accessing your sensitive data, the Library will apply the legal rules and provide limited access to the sensitive data putting password on access. Or, you can publish only a description (i.e. the metadata) of your data without making the data itself openly accessible, or you can anonymize the sensitive data which enables you to place conditions around access to the data.

Further information related to sharing sensitive data can be asked to Research Data staff via openaccess@ku.edu.tr address of the Library.

Figure 1: Access control to open data, mediated/controlled access and closed data

e. KU Research Data Repository Services

Long-Term & Reliable Storage:

Koç University Research Data Repository (KU-RDR) offer a competitively alternative to storage on a reliable cloud, which can be costly to introduce and maintain, and to storage on external drives, where data may be physically vulnerable and there is little or no guarantee that lost, stolen or corrupted data can be recovered. The KU Repository provides long-term archival (for 10 years), open access and dissemination of the University's research outputs.

Data Preservation:

The term 'preservation' means ensuring something can still be seen or used over time. In the context of digital data, long-term preservation is the process of maintaining data over time so that they can still be found, understood, accessed, and used in the future.

You may think that by saving your data in one or more places you have made sure it is effectively preserved, but with digital technology developing so quickly, your digital data are at risk from one or more of the following problems:

- file formats might not be compatible with future software, and therefore unreadable
- even if a document can still be opened with new software, it may be altered to a degree as to no longer be understandable or reliable for continued research

- storage media may have been degraded, scratched or broken, especially if they are portable, such as USB sticks, so information might be lost
- the files or data will not be understood because there is no supporting documentation or metadata, or this has not been preserved correctly either.

Documentation and Metadata Support:

Suna Kıraç Library will provide the templates of essential documentation such as Data Management Plans, README files, Metadata Forms and Author Contract before you send your research data into the KU Research Data Repository (KU-RDR) by contacting with the data steward of the project.

Open access to your research publications and datasets:

Koç University believes that data should be open, accessible and reusable. Open access to knowledge (both datasets and research papers) inherently facilitates interdisciplinary research and pushes the boundaries of discovery.

Funders such as European Research Commision (ERC), Horizon 2020, Welcome Trust and TÜBİTAK are increasingly mandating good data practice, including data management plans and data sharing, and recognising the need for global collaboration on infrastructure and best practice. While not all research data can be open access, momentum is gathering to achieve a future where research data are widely Findable, Accessible, Interoperable and Reusable (FAIR).

Many researchers are motivated to share their data but are often face with challenges in doing so. To this credit, KU Research Data Repository (KU-RDR) offers KU Community logn-term data storing, preservation and sharing to the public, training and support related to research data to protect the instutional output beside to increase the visibility of research performance of the University and our academicians.

4. DATA ANALYSIS & VISUALIZATION TOOLS

a.R and Tableau,

R and Tableau, two of many tools used in data analysis and visualization. R is an open source statistical programming language consisting of packages that can aid in processing, analyzing, and visualizing data in the data life cycle. Tableau is a proprietary application known for its interactive dashboards and data stories used to visualize and share data.

R is an open-source statistical programming language that consists of packages for cleaning, analyzing, and visualizing data. Download and Install R: https://cran.r-project.org/

RStudio is the development environment used to write R commands.Download and Install RStudio: https://www.rstudio.com/products/rstudio/download/

To work with data in R, there are various packages you can use that perform a variety of functions. The official site to find packages in R is R CRAN (Comprehensive R Archived Network) site. A more user-friendly site to search and browse R packages is the METACRAN.

Subsetting is an important concept in R. It takes a portion or slice of the data we want to work with based on conditions we choose. After slicing the data, you can use a R package to help arrange and order the data.

TABLEAU: Tableau is a business intelligence application that can be accessed through a subscription or for free with Tableau Public. Dashboards in Tableau are a way to present data in a way that the user can interact with.

[Create an account on Tableau Public](#)

[Download and install the software](#)

b.Python and Jupyter Notebooks

Phyton (plotting with Matplotlib, simulating data, and visualizing the Mandelbrot Set) and **Jupyter Notebooks** are the other programs and platforms that you can analyze and visualize your data.

Python is a programming language. You can write instructions, i.e., code, in Python to carry out a specific set of tasks. Python is the work horse of this unit, and we will be using the language to create visualizations and conduct data analysis.

Jupyter Notebooks are the interface that we've used to interact with Python. These notebooks are incredibly useful for testing chunks of code and analyzing the output that those chunks of code yield.

5.SKL TRAINING & SUPPORT

A successful implementation of a data management policy and a strategy depends on their acceptance and adoption by all staff members. Training and awareness sessions facilitate the understanding of the different roles around data and provides a starting point for improving the data management practices.

You can contact openaccess@ku.edu.tr for your training and support requests on preparing data management plans. However, Suna Kıraç Library will be offering a range of both class and online trainings, webinars during the academic semesters to support the needs of Koç University students, researhers and faculty members in the topics below:

- KU Open Science Policy & Library as Open Access Point
- General introduction to research data management (RDM)
- FAIR principles in Research Data Management (RDM)
- Funders' requirements for data management and sharing
- Data Management Plan preparation
- Organising data folders and files
- Data backup and storage solutions: KU Research Data Repository (KU-RDR)
- How to use repositories for data sharing and searching for existing datasets
- Data ownership and licensing
- Working with confidential data (personally identifiable, commercially sensitive etc.)
- Using version control software
- Library Data Lab
- Data café informal drop in sessions