Practical archiving and Metadata guide: Yoda

This document outlines the current requirements for [archiving data](https://fgb-rdm.nl/rdm/definitions/Definitions.html#archivingdata) on [Yoda](https://fgb-rdm.nl/rdm/tools/YODAInstructionsFGB.html) (the default archive for archiving research data in FGB). More information on the general faculty archiving requirements is found in [this summary](https://fgb-rdm.nl/rdm/archiving/ArchivingGuidelinesSummary.html); this guide supplements that information with specific guidance on archiving in YODA.

Once the archiving process is finalized, all data should be removed from active storage locations and data collection and analysis software.

If you have any questions or would like to discuss your specific case with a Data Steward, please contact research.data.fgb@vu.nl.

The minimum requirements for archiving data in Yoda is to include:

1. [Apply basic data management](#r1)
2. [Include structured metadata](#r2)
3. [Include a ReadMe file](#r3)
4. [Include unstructured metadata (e.g. codebooks)](#r4)
5. [Include terms of use license](#r5)
6. [Include data access protocol](#r6)

Please find additional detailed information on Yoda here:

* <https://fgb-rdm.nl/rdm/tools/YODAInstructionsFGB.html>
* <https://yoda.vu.nl/site/getting-started/index.html>

**Must Read:**

Documentation and files will not be accessible in data packages that are either available upon request (restricted access) or archived for verification only (closed access). Therefore, you should identify the non-sensitive documentation and files you want to make openly available and archive them in a way that ensures they are accessible. There are two ways to do this.

1. Archive all data in Yoda with restricted or closed access metadata. Add all documentation and files you want to make available to the [Open Science Framework (OSF](https://osf.io/)). Then, obtain a DOI in OSF and link it in the ‘[related data packages (10c)’ section](https://yoda.vu.nl/site/getting-started/metadata-add.html). This allows users who come across your closed or restricted dataset to easily find and access the non-sensitive documentation and files.
2. In the root folder in Yoda, create one folder for all restricted research data and another for the documentation and files meant to be publicly accessible (see Fig. 1 for an example). Archive and publish these folders separately, making sure the [metadata](https://yoda.vu.nl/site/getting-started/metadata-add.html) clearly describes the contents of each folder.

fig.1

\* Openly accessible documents must not include any personal, confidential, or sensitive information. This guidance applies to all types of data packages. For restricted or closed access packages, certain documents may be shared across both the open and restricted/closed versions. ReadME files and licenses must be included with all submissions. The essential point is that all data, files, and documentation should be connected through a persistent identifier (DOI).

1. **Data management basics:**

Basic file management is the foundation for a successful deposit; data should align with the following requirements:

* [See also the initial planning process when setting up a YODA environment](https://fgb-rdm.nl/rdm/tools/YODAInstructionsFGB.html#Initial_planning)

|  |  |
| --- | --- |
| **Task** | **Completed** |
| Use consistent and meaningful filenames that reflect the file content, avoiding spaces and special characters; if data are sensitive or restricted, indicate this in the file name. |  |
| A standard folder structure has been applied which is logical and appropriately named.* NB: If you want some data to be open and some restricted, plan your folder structure accordingly. Access permissions are set at the folder level, so organize your data with this in mind.
* [Visit this resource for further information](https://fgb-rdm.nl/Images/yoda_archiving_instructions.pdf).
 |  |
| The file formats are accessible through Open-Source Software and are not reliant on proprietary software. * See examples of recommended formats at: <https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/>
* If converting data across file formats, check that no data or internal metadata have been lost or changed
 |  |
| Duplicate or redundant files have been removed. |  |

1. **Yoda Metadata form:**

The Yoda metadata form allows you to create structured and standardized metadata that describes the project you are archiving. For more information please see the [Yoda handbook](https://yoda.vu.nl/site/getting-started/metadata-add.html), here you will find information on the specific fields.

* See also the [tips for Archiving & Metadata in the FGB YODA Instructions](https://fgb-rdm.nl/rdm/tools/YODAInstructionsFGB.html#furtherArchiving)

|  |  |
| --- | --- |
| **Task** | **Completed** |
| Ensure all recommended and mandatory fields are filled in. |  |
| If applicable, fill in optional fields. |  |
| Check that contents are suitable using the metadata descriptions found here: <https://yoda.vu.nl/site/getting-started/metadata-add.html> |  |
| Fill in metadata form at the top level, then copy the metadata to all subfolders. |  |
| Adjust subfolder metadata once copied to reflect its contents. |  |
| Check the fields ‘Data Classification’, ‘License’ and ‘Data Package Access’ to ensure they reflect the desired access permissions. |  |
| If you wish to link other datasets, publications, or resources after a dataset has been archived then you can subsequently update the metadata form under ‘Related Data Package’. |  |

1. **ReadMe file:**

A ReadMe file is required with all data submissions, this document should include all the information necessary to understand and interpret the dataset.

- For more information on ReadMe files [visit this link and see point 9](https://fgb-rdm.nl/rdm/archiving/ArchivingGuidelinesFull.html#section2.1.1).

- An example ReadMe is available [here](https://fgb-rdm.nl/Images/readme_example.txt).

|  |  |
| --- | --- |
| **Task** | **Completed** |
| Use FGB template in MarkDown: <https://fgb-rdm.nl/rdm/archiving/README_template.md> |  |
| Check that the ReadMe file is filled in fully, along with any additional information which would be useful for verification or reuse in the future. |  |
| ReadMe file is submitted along with the data package under ‘Documentation’ and made publicly available. |  |

1. **Codebooks (if applicable):**

Codebooks or data dictionaries add context and insight into your data and make it interpretable and reusable by others.

* [For more information visit this resource on codebooks](https://fgb-rdm.nl/rdm/documentation/Codebooks.html).

|  |  |
| --- | --- |
| **Task** | **Completed** |
| Check that the Codebook/ data dictionary is filled in fully, along with any additional information which would be useful for variable interpretation or reuse in the future. |  |
| Codebook/ data dictionary file is submitted along with the data package under ‘Documentation’ and made publicly available. |  |

1. **License:**

Every dataset is required to have a license which describes the purposes for which it can be accessed. Some data packages will have a mixture of licenses as some subsections of the data may be accessible under differing conditions. If so, an overview of the data package with corresponding licenses should be included in the ReadMe file.

* There are a variety of licensing options available to researchers, you can download and read more about licenses [here](https://yoda.vu.nl/site/getting-started/selecting-license.html).

|  |  |
| --- | --- |
| **Task** | **Completed** |
| Select a license which corresponds to the access conditions of your data.* For more information on selecting a license, visit the [VU Licensing Guide](https://yoda.vu.nl/site/getting-started/selecting-license.html)
 |  |
| Include information on licensing in ReadMe file, along with corresponding files. |  |
| License is submitted along with the data package under ‘Documentation’ and made publicly available. |  |

1. **Data access protocol (if applicable):**

Conditions, restraints, and procedures on data reuse and access should be described in a data access protocol. This should be made openly available and ensures transparency and accessibility for data access requests.

|  |  |
| --- | --- |
| **Task** | **Completed** |
| Use FGB Data Access Protocol: |  |
| Data access protocol is completed with all necessary information |  |
| Data access protocol is submitted along with the data package under ‘Documentation’ and made publicly available.  |  |

**Example of structured vs unstructured metadata:**

|  |  |
| --- | --- |
| **Type** | **Example** |
| Structured Metadata  | Title: Youth Well-being Survey 2024Creator: Dr. Jan JaansonKeywords: Youth, mental health, education, well-beingDate created: 2024-01-15 |
| Unstructured data (codebook) | "This dataset includes responses from 500 adolescents (ages 12-18) on mental well-being, education, and social interactions. Variables include self-reported stress levels (1-10 scale), weekly study hours, and number of close friends. Missing values are marked as ‘NA’. Data collected via online survey." |