Innovation Growth Lab

Data Management Template

Data Template

Introduction	The Data Template is intended for use in Stage 2 of an experiment. The aim of this resource is to guide the collection and storage of the underlying datasets of the trials. The purpose of the template is to help you and your partners prepare your data collection and plan how you will structure and store the data. By thinking in advance about the data structure, you can better anticipate problems that might arise during collection. The template ensures your internal processes consistently track what data you are collecting and how, providing a record for transparency and quality control. It also aids the sharing of data among people who were not involved in the collection process by clarifying what each variable is and what values it can take. Finally, it facilitates the process of ultimately anonymising and archiving the data.
How to use this template	To use this template, download/open and save a copy for your personal use. Add content by clicking on the tables and writing text in the relevant sections.
Key terms	Project Summary: A brief overview of the project aims and its specific research questions. Scale of Data: An indication of the volume of data expected, including the number of observations and total participants. Mode of Data Collection: A description of how data is generated, such as through surveys, interviews, platform usage data, or administrative records. Data Quality and Standards: The processes used to ensure consistency across the trial, such as standardising criteria for different staff members collecting data. Metadata Table: A supplemental table describing the fields in the dataset, including coding for derived variables and permissions for data use.
Version	1.1 (23/02/2026)
	This template was adapted from [ORIGINAL SOURCE] (CITATION) with permission from the authors / which is an open access resource.

Data template form[3]

The following form captures the characteristics of the project and the intervention. This form should be presented together with the data (in the format described in Section 3 below), to provide the user a general understanding of the project. For instance, if you were to share the data with a statistician within your organisation, this form would provide a general background to explain the context in which the data was created.

Please fill in all of the required sections following the guidelines (you can delete the guidelines as you fill in the form). As different sections ask related questions, we encourage you to read all the sections before beginning to fill in the form. This will avoid the need to redraft certain sections. In most cases, you only need a short description.

1. Project

1.1. Project title
Guidance: enter the proposal name here.

1.1. Author and contact details
Guidance: enter the name of the author filling out this form, and their contact details (email address, telephone).

[3] The following form is in part adapted from the UKRI Template on DMPOnline, available at https://dmponline.dcc.ac.uk/public_templates

2. Description of the data

This section describes in detail the characteristics of the data to be collected.

2.1. Project summary
Guidance: please provide a brief summary of the project aims, and its research questions.

2.2. Intervention
Guidance: please describe the intervention – i.e. what your project will deliver for the project participants. Please describe what each group of participants (e.g. control and treatment groups) will get as part of the intervention.

2.3. Type of data
Guidance: describe the types of data that you plan to collect (qualitative, quantitative).

2.4. Format of the data
Guidance: please describe the format in which the data is stored (file format, software used).

2.5. Scale of the data
Guidance: please provide an indication of the scale of the data – number of observations you plan to collect, number of participants, etc.

3. Data collection and generation

This section describes how the data is collected or generated

3.1. Mode of data collection and/or generation
Guidance: please describe how the data will be collected and/or generated (e.g. surveys, interviews, administrative data, platform date).

3.2. Data quality and standards
Guidance: please describe how the consistency and quality of data collection/generation will be controlled and documented (for instance, if different people collect data, how will you ensure they use the same standards and criteria?).

3.3. Dealing with missing values
Guidance: please describe how you plan to deal with missing values, e.g. values that you cannot collect because a firm does not respond to a survey.

3. Data table[4]

The data table below should be used to store individual firm/participant data. Each participant should be assigned a unique ID, with a row for each time data is collected and columns for variables including treatment(s) status (i.e. assigned to the treatment group or control); characteristics (e.g. size of firm); time period the data relates to; and outcome data (e.g. turnover or investment in new product development). Please note that you might have cells for which you do not collect data - e.g. because a firm failed to respond to a survey; in those cases, please specify in the metadata table (see below), how you plan to deal with these missing values (e.g. random imputation). Depending on your data collection plan, you might initially have a number of data tables, each originating from a different data source (e.g. one for baseline data collected at the application stage; another one for survey responses; a third one for administrative data). Nonetheless, if you plan to match data from different sources in your analyses, you will likely need to collect them all in one table.

Field name	Type	Description
ID	Mandatory	Unique ID to identify the trial participant (e.g. firm, entrepreneur). Several rows with data may be collected for the same participant; all should specify the ID of the participant they relate to.
Treatment	Mandatory	Treatment group for the participant. (If more than one treatment group, identify which one).
Baseline outcomes	Optional	If baseline data is collected for the outcome(s), please provide the values (specify in the metadata table the timeline for the baseline). Please note you may have collected data on more than one outcome at baseline.
Follow-up outcomes	Mandatory	The value taken by the outcome(s) at each follow up. Please describe in the metadata table the timeline for the follow-up(s)). Please note you may have collected data on more than one outcome at follow up.
Background characteristics	Optional	The value of any background characteristic for which you may collect data (please specify in the metadata table the details). Please note you may have collected data on more than one background characteristic at baseline.
Cluster ID	Optional	If randomising or analysing the data by cluster, please specify the cluster name (please describe the details in the metadata table).

[4] This data table is adapted from the one used by the Education Endowment Foundation (EEF), available at https://educationendowmentfoundation.org.uk/public/files/Evaluation/Submitting_your_data_to_the_FFT_archive/EEF_DataCollection_Specification.pdf

4. Metadata table[5]

The metadata gives an overview of the fields and items collected in the data table above. Moreover, it describes the codes used to characterise certain values (e.g. if a value is on a Likert scale from 1 to 5, it describes the scale and what it means). This will help anyone who is not familiar with the data collection process to understand what the data means. For instance, if an analyst from a partner organisation were to receive the raw data in a table from Section 2 above, will they be able to understand the names of the variables, or the values they take (such as 1 vs 0 for 'treatment')? This metadata table should be completed to ensure that they do.

For each of the items (rows) described in the data table above, insert a row describing them. In particular, describe its name (e.g. 'participant ID'), whether it is optional or mandatory, its 'code frame' (that is, what values it might take in your data - e.g. for a Likert scale, 1 through 5), and a description detailing what this means. (See an example metadata table in the Annex).

Field name	Mandatory	Code frame	Description
Item_name	Y	N	The corresponding field name in the data table
Item_description	Y	N	A label or description of the field name in the data table
Item_code	N	N	The code used in the value/cell of the field in the data table
Item_code_label	N	N	The label or description for the coded value.

[5] This metadata table is adapted from the one used by the Education Endowment Foundation (EEF), available at https://educationendowmentfoundation.org.uk/public/files/Evaluation/Submitting_your_data_to_the_FFT_archive/EEF_DataCollection_Specification.pdf