**Metadata template notes**

*Scientific Data* provides full structured metadata records (in ISA format; www.isa-tools.org) to document the provenance, manner of generation, and location in public repositories of the datasets linked to our Data Descriptors. Capturing metadata in a standardised way will facilitate search and retrieval of your data for future use, and therefore increase the likelihood your data will be used and cited.

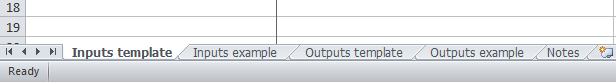
In order to generate these records, we require information about the initial inputs (e.g. samples, subjects) that were used in your study and the methods that were applied to these inputs to produce your data outputs. The Excel template in this zip package will help you to provide this information. This document is designed to be read in conjunction with the attached Excel file – it provides an overview and examples of how different data collection processes can be captured in the template.

The completed template (in xlsx fomat) should be uploaded to the manuscript submission system as ‘Data Descriptor Worksheet’. The *Scientific Data* curation team will use the tables in the template to create the final metadata record that will accompany your published Data Descriptor article.

Metadata records from previously published Data Descriptors can be browsed at <http://scientificdata.isa-explorer.org/>

**Template structure**

The template guides the generation of two types of table, one to describe the **Inputs** of your study and one to describe the data **Outputs** of your study. Each Data Descriptor manuscript will have only one **Inputs** table, and one or more **Outputs** table(s). These tables can be found in the Inputs template and Outputs template tabs of the associated Excel file, while illustrative examples are found in Inputs example and Outputs example tabs.



Bear in mind that the final metadata record will complement your Data Descriptor publication. To this end, the **Inputs** and **Outputs** tables should record the experimental workflow by which the described data outputs were generated, and the Methods declared in the completed template should match the Methods sub-headings in your manuscript.

The completed template should *not* contain the actual data outputs (only their description). All data outputs should be archived to the appropriate repository and *cited* in the Data Descriptor manuscript.

DATA FROM HUMAN STUDY PARTICIPANTS

If your data is derived, in part or in full, from human participants e.g. clinical data, survey data, please consult our **sensitive data checklist** to ensure that the metadata you provide does not contain any direct identifiers or more than three indirect identifiers.

***Inputs table:***

Describe your sources and samples (or inputs) in the **Inputs** template tab of the Excel template. This table should include details of how sources and samples were acquired, generated and any other aspects about the study inputs that potential data users would need to know.

* Each Data Descriptor should have only one **Inputs** table. It should include all of your samples.
* The **Inputs** table should *not* contain your data. It should focus on the sources for the samples that were used to generate the data outputs.
* The methods declared in the **Inputs** table should correspond to the Method subheadings in your Data Descriptor.

If this information is available in table form in your manuscript or at the repository, you can simply direct us to its location.

The examples below show how different sample generation workflows can be captured in the **Inputs** table.

Example 1: The same Source (with specified Characteristics) was subjected to Method 1, followed by Method 2 to create Samples 1 and 2 (with specified Characteristics)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Characteristics** | **Method** | **Method** | **Sample Name** | **Characteristics** |
| Source | Source characteristic | Source entity processing method 1 | Source entity processing method 2 | Sample 1 | Sample characteristic |
| Source | Source characteristic | Source entity processing method 1 | Source entity processing method 2 | Sample 2 | Sample characteristic |

Example 2: The same Source was subjected to Method 1 OR Method 2 to create Samples 1 and 2, respectively

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source Name** | **Characteristics** | **Method** | **Sample Name** | **Characteristics** |
| Source | Source characteristic | Source entity processing method 1 | Sample 1 | Sample characteristic |
| Source | Source characteristic | Source entity processing method 2 | Sample 2 | Sample characteristic |

Example 3: Multiple Sources were subjected to Method 1 followed by Method 2 to create Samples 1, 2 and 3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Characteristics** | **Method** | **Method** | **Sample Name** | **Characteristics** |
| Source 1 | Source characteristic | Source entity processing method 1 | Source entity processing method 2 | Sample 1 | Sample characteristic |
| Source 1 | Source characteristic | Source entity processing method 1 | Source entity processing method 2 | Sample 2 | Sample characteristic |
| Source 2 | Source characteristic | Source entity processing method 1 | Source entity processing method 2 | Sample 3 | Sample characteristic |

***Outputs table:***

Use the **Outputs** template tab(s) to describe the how the samples (the same as declared in the Inputs table) were processed to generate the data outputs. **Outputs** tables should include details of how the samples were processed to create the data that are cited in the Data Citations section of your Data Descriptor manuscript.

* Each Data Descriptor may have more than one **Outputs** table.
* The **Outputs** table should *not* contain your data. It should focus on the workflow by which your samples were used to generate the data outputs.
* The methods declared in the **Outputs** table should correspond to the Method subheadings in your Data Descriptor.

The examples below show how different data output generation workflows can be captured using the **Outputs** table.

Example 1: One Sample is used to generate each Data output.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample Name** | **Method** | **Method** | **Data output name** | **Raw Data File** | **Data Repository** | **Data Record Accession** | **Data Record URI** |
| Sample 1 | Sample processing method 1 | Sample processing method 2 | Output name 1 | Data file name 1 | Data repository name | Citable data reference | URL to access the data |
| Sample 2 | Sample processing method 1 | Sample processing method 2 | Output name 2 | Data file name 2 | Data repository name | Citable data reference | URL to access the data |

Example 2: One Sample is used to generate distinct Data outputs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample Name** | **Method** | **Data output name** | **Raw Data File** | **Data Repository** | **Data Record Accession** | **Data Record URI** |
| Sample 1 | Sample processing method | Output name 1 | Data file name | Data repository name | Citable data reference | URL to access the data |
| Sample 1 | Sample processing method | Output name 2 | Data file name | Data repository name | Citable data reference | URL to access the data |
| Sample 2 | Sample processing method | Output name 3 | Data file name | Data repository name | Citable data reference archive | URL to access the data |

Example 3: Distinct Methods are used on the same (or different) Samples to generate distinct types of data.

If distinct protocols were used to create different types of data from your samples, create a distinct Outputs tab in the Excel file for each type of data output (see image). These different types of data outputs may be archived to the same or different repositories.



Outputs\_template\_1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample Name** | **Method** | **Data output name** | **Raw Data File** | **Data Repository** | **Data Record Accession** | **Data Record URI** |
| Sample 1 | Sample processing method 1 | Output name 1 | Data file name 1 | Data repository 1 | Citable data reference | URL to access the data |
| Sample 3 | Sample processing method 1 | Output name 2 | Data file name 1 | Data repository 1 | Citable data reference | URL to access the data |

Outputs\_template\_2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample Name** | **Method** | **Method** | **Data output name** | **Raw Data File** | **Data Repository** | **Data Record Accession** | **Data Record URI** |
| Sample 1 | Sample processing method 2 | Sample processing method 3 | Output name 3 | Data file name 2 | Data repository 2 | Citable data reference | URL to access the data |
| Sample 2 | Sample processing method 2 | Sample processing method 3 | Output name 4 | Data file name 2 | Data repository 2 | Citable data reference | URL to access the data |

Example 4: The initial data outputs are processed to derive a second set of data outputs.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample Name** | **Method** | **Data output name** | **Raw Data File** | **Data Repository** | **Data Record Accession** | **Data Record URI** | **Method** | **Derived Data File** | **Data Repository** | **Data Record Accession** | **Data Record URI** |
| Sample 1 | Sample processing method 1 | Output name 1 | Data file name 1 | Data Repository Name | Unique identifier for the data archive | URL to access the data archive | Sample processing method 2 | Data file name 2 | Data Repository Name | Unique identifier for the data archive | URL to access the data archive |

Duplicate the Derived Data columns to the right of the table as many times as required to capture each subsequent Derived Data Output whose generation is described in the Data Descriptor. Data generated as part of the technical validation process do not need to be captured as Derived Data in the metadata tables.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data Record Accession** | **Data Record URI** | **Method** | **Derived Data File** | **Data Repository** | **Data Record Accession** | **Data Record URI** | **Method** | **Derived Data File** | **Data Repository** | **Data Record Accession** | **Data Record URI** |
| Unique identifier for the data archive | URL to access the data archive | Sample processing method 2 | Data file name 2 | Data Repository Name | Unique identifier for the data archive | URL to access the data archive | Sample processing method 2 | Data file name 2 | Data Repository Name | Unique identifier for the data archive | URL to access the data archive |

***Feedback:*** Please send any feedback on this metadata template to [scientificdata@nature.com](mailto:scientificdata@nature.com), marked for attention of the Data Curation Editor.