

# ADDIS excel/table study data interchange format

J.J. de Keijser, D.A.P. Reid

March 7, 2019  
v1.0

## 1 Summary

This document describes the required structure for an Excel file to be a valid document for import to the ADDIS study data repository. The format aims to be a compromise between human- and machine-readability.

We first discuss the single-study file structure, followed by explanation of the extra features of the dataset (i.e. multi-study) format in section 3.

Fixed headers are shown in **bold**.

References are shown in *italic* e.g. ‘=Concepts!B2’.

## 2 ADDIS study export

An example Excel study file is available at <https://drugis.org/files/excelStudyExample.xlsx>. We recommend keeping it open while reading this document for reference. Alternatively, you can create your own export example by clicking the export button for any study or dataset in the ADDIS web application. The following subsections each explain one sheet of the document.

### 2.1 Study Data

The Study Data sheet stores meta-data about the study, its arm structure, and the measurements that were reported for each variable and measurement moment.

Since this sheet is large and complex, we explain it block by block, with each ‘block’ corresponding to one of the headers in the first row of the sheet. We discuss the block moving from left to right.

#### 2.1.1 Study Information

The **Study information** block contains global information about the study. It is laid out as in table 1. Note that each cell in the bottom data row vertically spans all the arms, meaning it is as tall as the number of arms in the study. The **ID** and **addis url** cells show the source of the data if the file was exported, and are ignored when importing. They can be left blank when creating a new file.

Table 1: Study information

Study Information							
<b>ID</b>	<b>addis url</b>	<b>title</b>	<b>group allocation</b>	<b>blinding</b>	<b>status</b>	<b>number of centers</b>	<b>objective</b>
text*	URL*	text	text	text	text	number	text

#### 2.1.2 Population information

The **Population information** block contains the indication and eligibility criteria of the study. It is laid out as in table 2. Note that the indication information cell in this row spans all the arms, meaning it is as tall as the number of arms in the study.

Table 2: Population information

Population information	
<b>indication</b>	<b>eligibility criteria</b>
text	text

### 2.1.3 Arm information

The **Arm information** block contains the title and description of each arm in the study. It is laid out as in table 3. Note that there is one data row per arm, and that there is always a ‘Overall population’ group as well.

Table 3: Arm information

Arm information	
<b>title</b>	<b>description</b>
text (arm 1)	text
...	...
Overall population	

### 2.1.4 Measurement data

The **Measurement data** block contains the measured values for each variable measured in the study. This block is the most complex, with several nested block structures. A full layout can be found in table 9. We discuss the composing blocks below.

The measurement data block contains one variable block for each variable, as shown in table 4. Variable blocks are added horizontally (indicated with the ... column).

Table 4: Measurement data

Measurement data		
variable block 1	...	variable block n

### 2.1.5 Variable block

A variable block is laid out as in table 5, with a *variable name* cell spanning the block horizontally, (a reference to the name as entered on the ‘Concepts’ sheet). Below each *variable name* cell, there is the **variable type** header, **measurement type** header, and any number of measurement moment blocks: one per measurement moment at which the variable was measured. Valid values for the variable type are shown in table 6. Valid values for the measurement type are shown in table 7.

Table 5: Variable block

variable type	measurement type			
enum	enum	measurement moment block 1	...	measurement moment block n

Table 6: Variable types

baselineCharacteristic
endpoint
adverseEvent

### 2.1.6 Measurement moment block

A measurement moment block is laid out as in table 8. It consists of a **measurement moment** header, followed by one header per measured result property. Under the **measurement moment** header is the

Table 7: Measurement types

dichotomous
continuous
survival

*measurement moment* reference (referring to the ‘Measurement moments’ worksheet), which vertically spans all arm rows. Below each result property header comes one cell per arm, with for each cell the value (if any) measured in the study for that combination of arm, measurement moment and result property.

Table 8: Measurement moment block

measurement moment	result property 1	...	result property n
reference	value	...	value
	value	...	value
	value	...	value

Table 9: Full measurement data

Measurement data						
<i>variable name</i>						...
variable type	measurement type	measurement moment	result property 1	...	result property n	...
text	text	text	value	...	value	...
			...	...	...	...
			value	...	value	...

## 2.2 Activities

The **Activities** worksheet stores the actions that were performed in the study. Its most basic layout is shown in table 10. The possible values for the activity type enum are shown in table 11.

If any of the activities in the table are a ‘drug treatment’ activity, the table headers should be extended to cover this extra information: for each drug administered, there are the headers **drug label**, **dose type**, **dose**, **max dose**, **unit** and **periodicity**. An example with a drug treatment activity is shown in table 13. In the drug treatment’s row, the *drug label* cell is a reference to the drug’s label on the ‘Concepts’ worksheet. The possible values for the **dose type** column are shown in table 12. The *unit* cell is a reference to the unit’s label on the ‘Concepts’ worksheet. The **periodicity** cell indicates the interval between each dosage application. The interval’s duration is specified in ISO.8601 duration format.

Table 10: Activities simple

id	title	type	description
URI (activity 1)	text	enum	text
...	...	...	...
URI (activity n)	text	enum	text

Table 11: Activity types

screening
wash out
randomization
drug treatment
follow up
other

## 2.3 Epochs

The study’s temporal structure is described in the **Epochs** worksheet. Its layout is shown in table 14. An epoch’s **duration** is expressed in ISO.8601 duration format. Epochs may be instantaneous, in which

Table 12: Dosage types

fixed
titrated

Table 13: Activities complex

id	title	type	description	drug label	dose type	dose	max dose	unit	periodicity	...
URI (activity 1)	text	enum	text							...
...	...	...	...	...	...	...	...	...	...	...
URI (activity n)	text	enum	text	<i>drug label</i>	enum	value	value	<i>unit</i>	duration	...

case the duration is PT0S.

Table 14: Epochs

id	name	description	duration	isPrimary
URI (epoch 1)	text	text	duration	boolean
...	...	...	...	...
URI (epoch n)	text	text	duration	boolean

## 2.4 Study design

The **Study design** worksheet describes for each arm at which time which activities are performed. Its layout is shown in table 15. Note that since it only details how the arms, epochs and activities are linked, and these things are all defined elsewhere, this worksheet consists entirely of references. The *arm* references refer to the study data sheet's arm names, and the *epochs* and *activities* refer to their respective worksheets.

Table 15: Study design

arm	<i>arm 1</i>	...	<i>arm n</i>
<i>arm 1</i>	<i>activity</i>	<i>activity</i>	<i>activity</i>
...	<i>activity</i>	<i>activity</i>	<i>activity</i>
<i>arm m</i>	<i>activity</i>	<i>activity</i>	<i>activity</i>

## 2.5 Measurement moments

The **Measurement moment** worksheet lists the points in time at which measurements are taken. Its layout is shown in table 16. A measurement moment is defined as a duration relative to either the start or the end of an epoch. The *epoch* reference is to the epoch label on the **Epochs** worksheet. Valid values for **from** cells are shown in table 17. An measurement moment's **offset** is expressed in ISO\_8601 duration format.

Table 16: Measurement moments

id	name	epoch	from	offset
URI	text	<i>epoch</i>	enum	duration
...	...	...	...	...
URI	text	<i>epoch</i>	enum	duration

Table 17: Measurement moment epoch reference points

start
end

## 2.6 Concepts

The **Concepts** lists all the concepts (variables, drugs, units) which are used within the study. Its layout is shown in table 18. The valid values for the **type** column are shown in table 19. The **dataset concept uri** and **multiplier** columns are only relevant for dataset exports (see section 3.2) and can be left blank when creating a new study file.

Table 18: Concepts

id	label	type	dataset concept uri	multiplier
URI	text	enum	URI	value
...	...	...	...	...
URI	text	enum	URI	value

Table 19: Concept types

baseline characteristic
outcome
adverse event
drug
unit

## 3 Dataset export

A dataset is a collection of studies, along with a list of dataset-level concepts. Thus, the dataset export file has the same structure as that of the study export, but contains data for multiple studies on each worksheet. Further, there are two extra worksheets with additional dataset information, **Dataset information** and **Dataset concepts**. On all the study sheets, each original sheet, each study is separated by a row of white space. In all but the **Study data** sheet, the data for a study is preceded by a *reference* to the id of the corresponding study.

In order to handle studies that measure difference variable we let each study have its own variables. We currently do not harmonise across the entire dataset. In the future options such as concept selection might be considered.

### 3.1 Dataset information

The **Dataset information** worksheet contains meta-data about the dataset. Its structure is shown in table 20. The **ADDIS url** cell shows where the study came from if it is imported, and can be left blank when creating a new file.

Table 20: Dataset information

title	ADDIS url	description
text	URL	text

### 3.2 Dataset concepts

The **Dataset concepts** worksheet lists all the concepts defined at the dataset level. Its layout is shown in table 21. These dataset-level concepts are used to harmonise the study-level concepts, e.g. to indicate that the 'headache' concept from study A and the 'mild headache' concept from study B both refer to the same concept, and dataset-level analyses concerning headaches should include measurements from both studies.

On the **Concepts** sheet, for each study concept that should be linked, the **dataset concept** cell should contain a reference to an id cell on the **Dataset concepts** sheet. In the case of a mapping between unit concepts, there should also be a multiplier value, indicating how the study unit relates to the dataset unit. For example, when defining milligram in a study, it should refer to the 'gram' dataset concept, with

a multiplier value of 0.001. Supported multiplier values are those corresponding to the metric prefixes nano–mega, as shown in 22

Table 21: Dataset concepts

<b>id</b>	<b>label</b>	<b>type</b>
URI	text	text
...	...	...
URI	text	text

Table 22: Unit multipliers

Metric prefix	Value
mega	1000000
kilo	1000
hecto	100
deca	10
(none)	1
deci	0.1
centi	0.01
milli	0.001
micro	0.000001
nano	0.000000001