

Caring for data



Worksheet section	Contents
1	Caring for datasets
2	Create a data dictionary

Version: 1.0

This lesson has been created by effini in partnership with Data Education in Schools and Skills Development Scotland.

© 2022. This work is licensed under a [CC BY-NC-SA 4.0 license](#).



You are free to:

Share – copy and redistribute the material in any medium or format

Adapt – remix, transform and build upon the material

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.

NonCommercial — You may not use the material for [commercial purposes](#).

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

If you require this document in an alternative format, such as large print or a coloured background, please contact

hello@effini.com

or

**4th Floor, The Bayes Centre
47 Potterrow
Edinburgh
EH8 9BT**

1. Caring for datasets

Section 1.1 (recall)

1) Fill in the missing gaps in the different types of datasets

Metadata Data about the

Reference data Data used by other data sources such as

Main data Dataset that contains the information that is important to an organisation.

Transactional data Data that records normally with datetime information

Section 1.2 (rephrase)

2) Why is important to care for data?

3) Why does reference data need to be cared for the most?

Section 1.3 (apply)

For each of these datasets state the type of dataset and describe why.

4)

planet	type	size_km
Mercury	Terrestrial	2,440
Venus	Terrestrial	6,052
Earth	Terrestrial	6,371
Mars	Terrestrial	3,390
Jupiter	Gas giant	69,911
Saturn	Gas giant	58,232
Uranus	Ice giant	25,362
Neptune	Ice giant	24,622

Dataset type

Reason for choice

5)

patient_id	name	DOB
AQ1245	F. Copper	25/08/1998
QW1254	M. Reid	19/12/1998
ER7854	L. Power	12/02/2001
FT6954	T. Young	22/10/2001
GT3564	D. Whyte	23/03/2004
PL1289	M. Trent	22/09/2005

1. Caring for datasets

Dataset type

Reason for choice

6)

date	number_visitors
05/04/2023	5124
06/04/2023	6475
07/04/2023	4578
08/04/2023	210
09/04/2023	1234
10/04/2023	2546

Dataset type

Reason for choice

7)

Dataset name:	Public Toilets and Comfort Schemes
Description:	The location of Public toilets and comfort schemes within the Perth and Kinross area.
Maintainer:	Perth & Kinross Council
Created:	18-Aug-22
Last updated:	12-Sep-22
Language:	English
Source of data:	https://data.pkc.gov.uk/dataset/public-toilets-and-comfort-schemes

Dataset type

Reason for choice

Section 1.4 (active)

- 8) Public Health Scotland shares open data for the public to use. Using their website (link below) find an example of reference dataset. Fill in the name and description of the dataset below.

<https://www.opendata.nhs.scot/dataset>

Dataset name:

Description:

- 9) What would happen if this reference data was not cared for?

2. Create a data dictionary

Section 2.1 (recall)

1) Why are data dictionaries important?

2) What are the 3 pieces of information about a variable should you include in a data dictionary?

Section 2.2 (apply)

For each of these datasets, fill in the gaps for the associated data dictionary.

3)

planet	type	size_km
Mercury	Terrestrial	2,440
Venus	Terrestrial	6,052
Earth	Terrestrial	6,371
Mars	Terrestrial	3,390
Jupiter	Gas giant	69,911
Saturn	Gas giant	58,232
Uranus	Ice giant	25,362
Neptune	Ice giant	24,622

Data dictionary

variable_name		type
planet	Name of the plant in our solar system	string
type	The type of the planet. Either terrestrial, gas giant or ice giant.	
	The diameter of the planet in km	integer

4)

ice_cream	price_per_scoop
Chocolate	1.20
Strawberry	1.20
Vanilla	1.00
Mint Choc	1.25
Raspberry	1.20
Irn Bru flavour	1.30

Data dictionary

variable_name	description	
	Flavour of ice cream available in a shop.	
	Price in £s of each scoop of ice cream	floating point

2. Create a data dictionary

5)

date	number_visitors
05/04/2023	5124
06/04/2023	6475
07/04/2023	4578
08/04/2023	210
09/04/2023	1234
10/04/2023	2546

Data dictionary

variable_name		type
	Date (DD/MM/YYYY)	datetime
	Number of visitors to the park per day	

6)

statues	height	location
Scott Monument	61	Edinburgh
Statue of Liberty	93	New York
Little Mermaid	1	Copenhagen
Christ the Redeemer	38	Rio de Janeiro
Nelson's column	52	London
The Kelpies	30	Falkirk

Data dictionary

	description	type
statues	Names of famous statues	
	Maximum height in m	
location	City/town of the statues	

7)

time	temperature
09:00	7
10:00	8
11:00	8
12:00	9
13:00	6
14:00	6
15:00	5
16:00	5
17:00	5
18:00	5

Data dictionary

	description	type
time	Time in hours of the measurements	
	Air temperature in degrees celcius	

2. Create a data dictionary

Section 2.3 (active)

- 8) In section 1 question 7, you found a dataset from the Public Health Scotland website. Fill in the data dictionary for your chosen dataset.

Dataset name:

Description:

Data dictionary

variable_name	description	type