What is Data?

Understand data, data type and data format

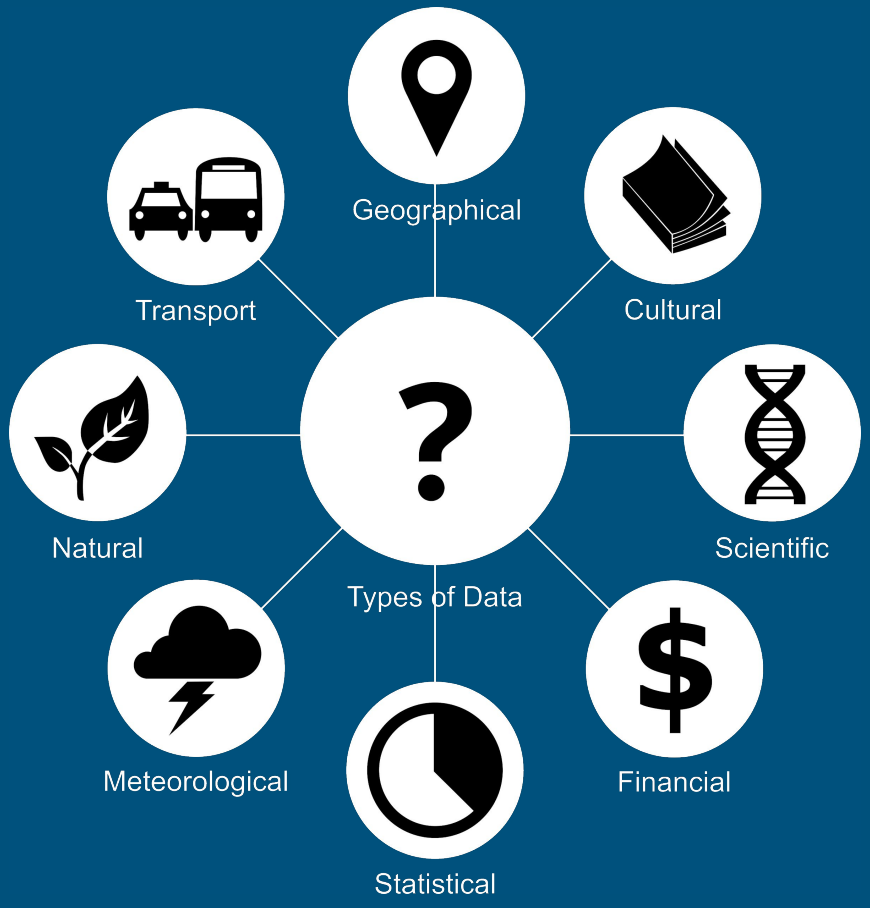
Data as a collection of fact

Data

Analysis

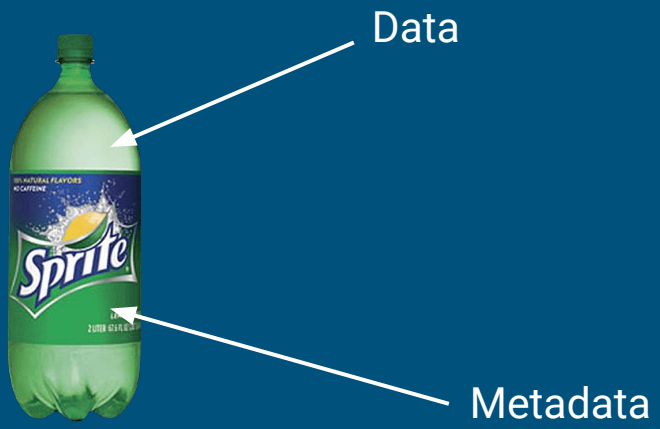
Information

What is data?


Geographical
Transport
Cultural
Natural
?
Scientific
Types of Data
Meteorological
Financial
Statistical

Metadata

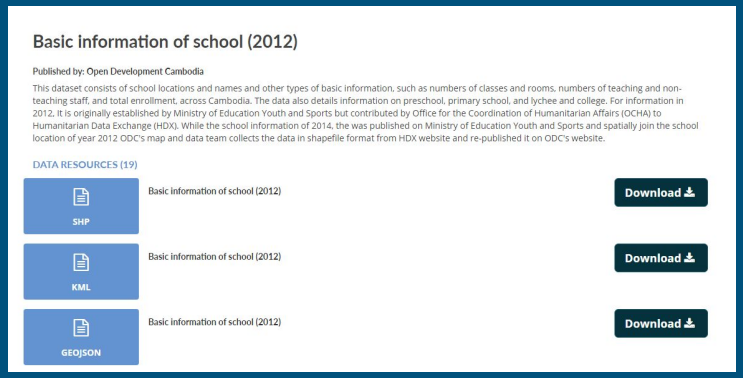
Metadata is data of data. It tell information about data.



Metadata should include:

- Title
- Producer
- Language
- Geography
- Limitation
- Date and time
- Accuracy
- Data format
- Use restriction

Metadata



Basic information of school (2012)

Published by: Open Development Cambodia

This dataset consists of school locations and names and other types of basic information, such as numbers of classes and rooms, numbers of teaching and non-teaching staff, and total enrollment, across Cambodia. The data also details information on preschool, primary school, and lychee and college. For information in 2012, it is originally established by Ministry of Education Youth and Sports but contributed by Office for the Coordination of Humanitarian Affairs (OCHA) to Humanitarian Data Exchange (HDX). While the school information of 2014, the was published on Ministry of Education Youth and Sports and spatially join the school location of year 2012 ODC's map and data team collects the data in shapefile format from HDX website and re-published it on ODC's website.

DATA RESOURCES (19)

SHP	Basic information of school (2012)	Download
KML	Basic information of school (2012)	Download
GEOJSON	Basic information of school (2012)	Download

Field	Value
Dataset topic category	• Access to education • Education and training • Pre school • Primary and secondary education
Language	• English
Use limitations	By accessing this website or database, users agree to take full responsibility for reliance on any site information provided and to hold harmless and waive any and all liability against individuals or entities associated with its development, form and content for any loss, harm or damage suffered as a result of its use.
GeoNames	• Cambodia
West bounding coordinates	102.34538725917648
East bounding coordinates	107.46499307467106
South bounding coordinates	10.438913769427716
North bounding coordinates	14.417659200338127
Spatial Reference System	WGS 84 / UTM zone 48N (EPSG:32648)
Positional Accuracy	There are no known issues with accuracy.
Logical Consistency	There are no known issues with logical consistency.
Completeness	There are no known issues with completeness.
Process Step	Basic Information of school 2012: Open Development Cambodia (ODC) downloaded the data from HDX (https://data.humdata.org) as ESRI shapefile and exported it to KML and GeoJSON extensions. Then, ODC visualized this data on CatoDB for creating a layer on ODC map explorer https://opendevelopmentcambodia.net/map-explorer. Basic Information of school 2014: ODC downloaded the non-spatial data from the Ministry of Education Youth and Sports website then spatially join with the locations from Basic Information of school 2012. Then ODC exported it to KML and GeoJSON extensions. Then, ODC visualized this data on CatoDB for creating a layer on ODC map explorer.

Qualitative and quantitative data

- Qualitative data is everything that refers to the quality of something: A description of colours, texture and feel of an object , a description of experiences, and interview are all qualitative data.
- Quantitative data is data that refers to a number. E.g. the number of students, the size, the cost, a score on a test etc.

Other ways to describe data

- Categorical data puts the item you are describing into a category: In our example the condition “public” would be categorical (with categories such as “private”, “public” , etc.)
- Discrete data is numerical data that has gaps in it: e.g. the count of footballs. There can only be whole numbers of footballs (there is no such thing as 0.3 test score).
- Continuous data is numerical data with a continuous range: e.g. size of the class can be any value or the time of study (as opposed to your class size, which is discrete): In continuous data, all values are possible.

Aggregate and disaggregate data

Disaggregate data

ID	Commune	School Name	Village	School Type	Class	Room	Location	Teaching Staff	Non-teaching Staff
1	Banteay Neang	Banteay Neang	Banteay Neang	Primary	17	9	Rural	21	3
2	Banteay Neang	Banteay Neang	Banteay Neang	Pre school	4	4	Urban	4	1
3	Banteay Neang	Trang	Trang	Primary	1	1	Rural	1	0
4	Banteay Neang	Pongror	Pongror	Primary	6	3	Rural	4	1
5	Banteay Neang	Pongror	Pongror	Pre school	1	1	Rural	1	0
6	Banteay Neang	Kauk Tunloap	Kauk Tunloap	Primary	6	4	Rural	5	2

Aggregate and disaggregate data

Aggregate data

School Type	Count
Pre school	2
Primary	4

No. of Class	Count
1-10	5
11-20	1

No. of Room	Count
1-10	6
11-20	0

Location	Count
Urban	1
Rural	5

No. of Teaching Staff	Count
1-10	5
11-20	0
21-30	1

No. of Non-teaching Staff	Count
1-10	5
11-20	0
21-30	1

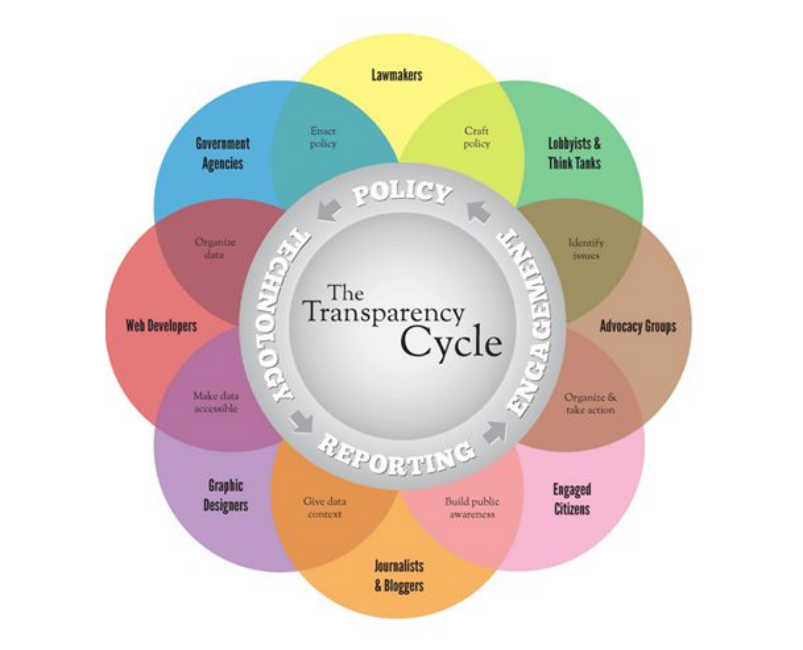
Open data

“Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike”. - *Open Definition*

Why data?

- Transparency and democratic control
- Participation
- Self-empowerment
- Improved or new private products and services
- Innovation
- Improved efficiency of government services
- Improved effectiveness of government services
- Impact measurement of policies
- New knowledge from combined data sources and patterns in large data volumes

Open data


Lawmakers
Enact policy
Craft policy
Government Agencies
Lobbyists & Think Tanks
POLICY
ENGAGEMENT
REPORTING
TECHNOLOGY
The Transparency Cycle
Organize data
Identify issues
Web Developers
Advocacy Groups
Make data accessible
Organize & take action
Graphic Designers
Give data context
Build public awareness
Engaged Citizens
Journalists & Bloggers