

[Context](#)

[Using the lexicon](#)

[Adding to the lexicon](#)

[History](#)

[Terms reviewed in previous cycles](#)

[Data that is closed](#)

[Data that is shared](#)

[Open data](#)

[Personal data](#)

[Big data](#)

[Data Spectrum](#)

[Internal Access](#)

[Named Access](#)

[Group-based Access](#)

[Public Access](#)

[Open API](#)

[Open standard](#)

[Data infrastructure](#)

[Terms newly open for comment in this cycle.](#)

[Machine-readable](#)

[Wishlist](#)

[Related Work](#)

Context

This is a public document where anyone can comment or make a suggestion to the current draft of the Open Data Institute's work on a data lexicon.

You can find more about the data lexicon in [this blog](#). The key section is this paragraph:

We are investigating whether we can openly grow a data lexicon to help solve this problem. Can we keep language around data accurate and useful while making it simpler so that more people can understand?

Whenever we simplify language we risk making a trade-off between accessibility/inclusion and accuracy as we seek to wide the debate so that more people can understand and join in.

The data lexicon is intended for use in multiple countries. Where possible definitions should not be bound to a specific legal or cultural context. In cases where legal accuracy is needed definitions should encourage people to refer to relevant legislation.

Different cultures, professions and purposes may require alternative or more specific definitions. In the lexicon we want to agree general definitions (i.e. ones suitable for creating a sufficiently common language for debate) whilst leaving space to also capture definitions that have been suitable for a specific purpose.

Using the lexicon

All content in this document is CC-???. Please link back to this source document if possible.

Adding to the lexicon

We encourage people to add comments after using the definitions and listening to any feedback that they receive.

History

On 2 October 2015 this document was updated to incorporate comments and some new draft definitions.

On 4 November 2015 this document was updated to incorporate additional comments and feedback from the Open Data Institute's [node network](#).

At the end of November 2015 we will re-update to incorporate comments and provide further draft descriptions for discussion.

All comments are publicly visible. If you wish to discuss any item in a different way than a comment then please email the team at policy@theodi.org or tweet [@peterkwells](#).

Terms reviewed in previous cycles

Definitions will be considered agreed unless significant comments emerge. Each definition has a standard definition for use in most debates/discussion along with a section on alternative definitions that people have found work for a particular purpose, context, group of professional groups.

Data that is closed

After debate about the difference between data that **should not** be shared - for security reasons, or because it's personal information - and data that is simply only available to an individual, to a team or within an organisation by default, we agreed on:

Data that can only be accessed by its subject, owner or holder.

Alternative definitions

Purpose	Definition	Contact (for more details)
<i>e.g. discussion with security experts</i>	<i>Closed data is data that can only be accessed by its subject, owner or holder. Passwords stored with a one-way encryption algorithm are a strong example of closed data. The password can only be accessed by its owner.</i>	<i>@peterkwells</i>

Data that is shared

This was trickier. Having agreed that access would be the anchor for variation across the spectrum, we settled on these 'shades' of access:

- **Named access** - "data that is shared only with named people or organisations"
- **Group-based access** - "data that available to specific groups who meet certain criteria"
- **Public access** - "data that is available to anyone under terms and conditions that are not 'open'"

Alternative definitions

Purpose	Definition	Contact
---------	------------	---------

		(for more details)
<i>e.g. discussion with public sector legal experts</i>	<i>Data sharing is when data is made accessible beyond its subject, owner and holder. The statutory definition of data sharing is in the 2014 Act XYZ.</i>	@peterkwells

Open data

Our definition of open data didn't change. Our one sentence definition continues to be:

Data that anyone can access, use and share.

For data to be considered 'open', it must be:

- **Accessible**, which usually means published on the web
- Available in a **machine-readable** format
- Have a **licence** that permits anyone to access, use and share it - commercially and non commercially.

A fuller and more precise definition is maintained by the [Open Definition](#) project.

Alternative definitions

Purpose	Definition	Contact (for more details)
<i>e.g. discussion with ABC</i>	<i>DEF</i>	@XYZ

Personal data

Data from which a person can be identified is personal data.

If data can be combined with other information to identify a person, that data will still be "personal data".

It's important to remember that personal data can be closed, or shared with specific people or organisations, or licensed as open data.

Nations may [have legislation](#) that explicitly defines personal data within a jurisdiction. Our general definition may define more or less data as personal data than the relevant legislation.

Alternative definitions

Purpose	Definition	Contact (for more details)
<i>e.g. discussion with ABC</i>	<i>DEF</i>	@XYZ
UK legal contexts	The Information Commissioner's Office defines personal data as "Personal data means data which relate to a living individual who can be identified – (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual." See: https://ico.org.uk/for-organisations/guide-to-data-protection/key-definitions/#personal-data	@peterkwells

Big data

A vendor-driven term often used to describe large quantities of rapidly changing data being collected from various sources.

There is no common definition of big data. Just as there is no common definition of the terms small data and medium data that appear in some literature.

It's important to remember that 'big data' can be closed, or shared with specific people or organisations, or licensed as open data.

Alternative definitions

Purpose	Definition	Contact (for more details)
<i>e.g. discussion with ABC</i>	<i>DEF</i>	@XYZ

Data Spectrum

Data exists on a [spectrum](#)

The spectrum ranges from closed to shared to open.

There are many shades on the spectrum. The following three terms are used to describe different types of data sharing on the [data spectrum](#). There are many other types of data sharing.

Internal Access

Lots of data might only be available within an organisation. Sometimes this is deliberate. Sometimes it simply doesn't have value for anyone outside the organisation.

When you describe data as 'internal access' you are explicitly saying that access to the data is limited inside your organisation or team.

Named Access

Sometimes data might be shared with specific people or organisations for a specific purpose. This will typically be explicitly assigned in a contract. Sometimes a data owner might give the holder of their data permission to share that data with specific groups.

For example: driving licence data might be shared between a driver and a company as part of the process for hiring a car.

Group-based Access

Sometimes data is available to special groups or people who meet certain criteria. Their access will be authenticated.

For example: health data might be made available to certain researchers - who have applied for access - for medical research purposes.

Alternative definitions

Purpose	Definition	Contact (for more details)
<i>e.g. discussion with ABC</i>	<i>DEF</i>	<i>@XYZ</i>

Public Access

Sometimes data is available to anyone, but under terms and conditions that are not open. The terms and conditions might only allow non commercial use, or restrict adaptations of the data. The licence limits the use.

For example: Twitter data is accessible to anyone who agree to a licence. The terms and conditions for its use are not open.

Alternative definitions

Purpose	Definition	Contact (for more details)
<i>e.g. discussion with ABC</i>	<i>DEF</i>	@XYZ

Open API

A means of accessing data based on an open standard: it is a public interface. The data accessed via an open API may be closed, shared or open data.

Alternative definitions

Purpose	Definition	Contact (for more details)
<i>e.g. discussion with ABC</i>	<i>DEF</i>	@XYZ

Open standard

An open standard is developed and maintained collaboratively and transparently, and can be accessed and used by anyone.

A more detailed definition can be found [here](#).

Alternative definitions

Purpose	Definition	Contact (for more details)
---------	------------	-------------------------------

e.g. discussion with ABC	DEF	@XYZ
--------------------------	-----	------

Data infrastructure

A data infrastructure consists of **data assets**, the organisations that **operate** and maintain them and **guides** describing how to use and manage the data. Strong data infrastructure is sustainably **funded** and has **oversight** that makes choices to maximise data use and value by meeting the needs of the **community**.

Alternative definitions

Purpose	Definition	Contact (for more details)
e.g. discussion with ABC	DEF	@XYZ

Terms newly open for comment in this cycle.

Machine-readable

Wishlist

More data terms for which our team have asked for clear and simple definitions to be added to the lexicon. Some of them are not commonly misused so much as simply unfamiliar to non technical audiences. If you want to add more items or suggest definitions please leave a comment:

API

- A promise by one service system to consistently provide a service data to another.

. . (e.g. **you** can contact **me** on this **phone number**)

Blockchain

Data

Dataset

Data subject

Data owner

Data holder

Linked data

Open source

Open access

Open government

Open identifiers

Exhaust data

Open city

Open innovation

Midata

Reference data

Administrative data

Code list

Data schema

Metadata

Distributed ledger

Related Work

This section is for links to other sites discussing definitions for language about data.

<http://opendefinition.org/>

<https://data.gov.uk/glossary>