What is 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.

The [full Open Definition](http://opendefinition.org/okd/) gives precise details as to what this means. To summarize the most important:

* **Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
* **Re-use and Redistribution:** the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.
* **Universal Participation:** everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, ‘non-commercial’ restrictions that would prevent ‘commercial’ use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

If you’re wondering why it is so important to be clear about what open means and why this definition is used, there’s a simple answer: **interoperability.**

Interoperability denotes the ability of diverse systems and organizations to work together (inter-operate). In this case, it is the ability to interoperate - or intermix - different datasets.

Interoperability is important because it allows for different components to work together. This ability to componentize and to ‘plug together’ components is essential to building large, complex systems. Without interoperability this becomes near impossible — as evidenced in the most famous myth of the Tower of Babel where the (in)ability to communicate (to interoperate) resulted in the complete breakdown of the tower-building effort.

We face a similar situation with regard to data. The core of a “commons” of data (or code) is that one piece of “open” material contained therein can be freely intermixed with other “open” material. This interoperability is absolutely key to realizing the main practical benefits of “openness”: the dramatically enhanced ability to combine different datasets together and thereby to develop more and better products and services (these benefits are discussed in more detail in the section on ‘why’ open data).

Providing a clear definition of openness ensures that when you get two open datasets from two different sources, you will be able to combine them together, and it ensures that we **avoid our own ‘tower of babel’: lots of datasets but little or no ability to combine them together into the larger systems where the real value lies.**

# What Data are You Talking About?

Readers have already seen examples of the sorts of data that are or may become open - and they will see more examples below. However, it will be useful to quickly outline what sorts of data are, or could be, open – and, equally importantly, what won’t be open.

The key point is that when opening up data, the focus is on non-personal data, that is, data which does not contain information about specific individuals.

Similarly, for some kinds of government data, national security restrictions may apply.