

Open Data: the Panton Principles



Published February 19, 2010

Citation

Willighagen, E. (2010, February 19). Open Data: the Panton Principles. *Chem-bla-ics*. <https://doi.org/10.59350/q5sed-jea02>

Keywords

Opendata

Copyright

Copyright © None 2010. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The [announcement](#) of the [Panton Principles is the big news today](#), though Peter already spoke about them [in May last year](#) (see coverage on [FriendFeed](#) and [Twitter](#)). The four principles list in their short versions:

- When publishing data make an explicit and robust statement of your wishes.
- Use a recognized waiver or license that is appropriate for data.
- If you want your data to be effectively used and added to by others it should be open as defined by the Open Knowledge/Data Definition – in particular non-commercial and other restrictive clauses should not be used.
- Explicit dedication of data underlying published science into the public domain via PDDL or CCZero is strongly recommended and ensures compliance with both the Science Commons Protocol for Implementing Open Access Data and the Open Knowledge/Data Definition.

I think these are very workable next steps in Open Date, perhaps even worthy end goals. [I endorse them](#).

Panton Principles

Principles for Open Data in Science

Principle 1: an explicit and robust statement

This is in my opinion the most important principle. Too often you find a database with really useful data, but without any clue about what you are allowed to do with this data. Of course, I can contact the authors, get their permission, etc. They probably like it that way, and I can even understand that. However, it does not scale, and it is slow. Even worse is the situation when the original composer gets missing in action. Both are equally valid, but explicit statements just make things easier.

Principle 2: use a waiver or license appropriate for data

This principle is debatable. Very much like the BSD-vs-GPL flamewars, some like copylefting, others do not. There is an important difference though. Software has the concept of interfaces, allowing to more easily share incompatible licenses cleanly separated by these interfaces. This, for example, allows you to run proprietary software on a Linux kernel. However, data sets do not have such a concept. There is not such thing as an interface between two numbers.

This makes the concept of mixing data sets different: because there is no such interface, any mixing can only happen between compatible licenses. This is one reason behind the choice of very liberal licenses like [CC0](#). This license, or waiver really, allows you to do anything, and most certainly, mix data sets.

And that makes things a lot easier. But then again, while these are noble goals, I rather see people use a copylefting licenses than no license at all.

Principle 3: non-commercial and other restrictive clauses should not be used

I think again making things easier is the goal. The non-commercial clause is interesting, and actually likely an important one. Consider course material, a course book. Those are commercial. Some even argued that many universities themselves are actually commercial entities.

Principle 4: the public domain via PDDL or CCZero is strongly recommended

I second these choices over a mere claim claim that the data is public domain. The PD concept has many meanings and not the same in every jurisdiction. In particular, differences between USA and EU law. Waiving these right, which is just the same as claiming public domain, works in any jurisdiction, again, making things a lot easier.

Open Data, Open Source, Open Standards are not goals

The underlying pattern of my comments must be clear: the principles make life easier. This is all what Open Source and Open Standards ([whatever those are](#)).

- ***The three pillars of the ODOSOS mantra is not goals, but merely the means of making life easier.***

The Panton Principles certainly make life easier in Open Data, and initiative like the [Linking Open Drug Data](#) in which I participate will greatly benefit from people adopting them.

The Principles do not solve all problems. There is still a lot of 'Open Data' licensed with unrecommended licenses. For example, the [NMRShiftDB](#) uses a GNU FDL license, and data from supplementary material of Open Access journal articles is like Creative Commons.

Is It Open Data?

[Home](#) [Make Enquiry](#) [Browse Existing](#) [About](#)

We've made it easy for you to make enquiries to data holder material they hold — and to record publicly the results of th

[How Does It Work?](#)

Or

[Browse Enquiries »](#)

Or

[Make an En](#)

Why Is Open Data Needed?

Another related initiative should certainly not go unnoticed either: [Is it Open Data?](#) is a service where you can try to resolve what the license is for one of those databases which is not quite Panton Principles compatible yet.

OK, one last thing. The [Dutch government is bursting](#), and I want to listen to the music. With permission, I have been hacking the Panton Principles endorsement page, and injected some

chem-bla-ics

extra span elements, to make it easier to machine process (again, to make things easier), so you can use the following one-liner to calculate the number of people endorsing the principles:

```
$ wget -O endorsed.html http://pantonprinciples.org/endorsed.html ; xpath -q -e "//span[@class='signature']/span[@class='Country']/text()" endorsed.html | sort | uniq -c
```

The current count is [hitting 44 now](#), and has not quite reached the [500 I had hoped for](#) yet:

```
1 Australia
1 Canada
1 Catalonia
2 Espana
2 France
6 Germany
1 Greece
1 Italy
1 Netherlands
1 New Zealand
1 Norway
1 Poland
1 Slovenia
1 Sweden
1 Switzerland
1 The Netherlands
9 UK
1 U.K.
1 United Kingdom
1 United States of America
9 USA
```

Anyone knows how we can convert this into some nice world map graphics with a few lines of code?

Now, I am looking for a bar in Uppsala to write up some ideas about what specifications are :)