Can That Data Be Shared?

The Panton Principles seeks to promote "open data" commons in science.

By [David Bollier]

The proud authors of the Panton Principles, in front of the Panton Arms in Cambridge, UK.

One of the big problems in science is the proliferation of databases whose content is technically incompatible or legally proprietary in some fashion — and therefore unable to be used by others in their research. For years a number of smart, committed scientists, law scholars and techies have grappled with the problem of making data accessible and re-useable. Now they have released a blueprint for doing so.

The [Panton Principles for Open Data in Science] is a major effort to articulate a clear definition of "open data" and help scientists make the right choices in trying to make their data "open." The principles set forth the general steps that scientists should take to create more effective and sustainable data commons.

The preamble to the Panton Principles reads:

Science is based on building on, reusing and openly criticizing the published body of scientific knowledge. For science to effectively function, and for society to reap the full benefits from scientific endeavors, it is crucial that science data be made open.

By open data in science we mean that it is freely available on the public internet permitting any user to download, copy, analyze, re-process, pass them to software or use them for any other purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. To this end data related to published science should be explicitly placed in the public domain.

The principles are important because they help sweep aside some dangerous misconceptions about open data. For example, the use of copyright-based licenses such as Creative Commons licenses won't work when applied to data. Data are not the same as other works, and so the practical value of copyright protection when applied to data is questionable.

For example, it's confusing whether copyright protection applies only to the data itself, to the database model (the structure and organization of the data), or to the data entry and output sheet. So if a CC

license were applied to a database, confusion is likely about what is covered by the license and what isn't. Then there is the core problem of whether facts — which normally cannot be copyrighted — would be eligible for copyright protection in databases.

The Panton Principles urges database owners to take four basic steps to help assure that their data can be shared:

1. When publishing data make an explicit and robust statement of your wishes.

2. Use a recognized waiver or license that is appropriate for data.

3. 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.

4. 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.

The Panton Principles were first drafted in July 2009 by Peter Murray-Rust, a chemistry professor at the University of Cambridge, England; Cameron Neylon of the Science and Technology Facilities Council (UK), Rufus Pollock of the Open Knowledge Foundation (UK) and University of Cambridge; and John Wilbanks of Science Commons, an offshoot of Creative Commons.

The naming of the principles has a quaint British backstory: it comes from the Panton Arms on Panton Street in Cambridge, England, which is not far from Peter Murray-Rust's chemistry lab. The first draft was written there, presumably over a series of beers, and later refined with the help of the members of the Open Knowledge Foundation Working Group on Open Data in Science.

So, if you've got data and want to make sure that they can be used and re-used to advance human knowledge, check out the Panton Principles. Make sure you are in fact making your datasets legally and technically accessible! You can endorse the Panton Principles at here.

Posted February 23, 2010