Open Data & The Panton Principles

John Dupuis

January 31, 2013

Ontario Library Association Super Conference

http://bit.ly/DupuisSC13

http://bit.ly/dupuisdatalinks

Panton Principles for Open Data in Science

- Drafted in July 2009 by Peter Murray-Rust, Cameron Neylon, Rufus Pollock & John Wilbanks
- Refined with help from the Open Knowledge Working Group on Open Data in Science
- Launched in February 2010
- Text is on the website

Ever wonder why they're called the Panton Principles?

pic

What open data is about

- Making data from public institutions and publicly funded research available using as open a license as possible
- Replicability
- New questions from existing data
- Combining data sets

What open data is NOT about

- Getting scooped
- Being denied credit/citation
- Data being misused or misunderstood
- Releasing data that shouldn't be public due to privacy concerns, etc.
- Releasing articles or other research outputs under the same or similar licenses

Biggest roadblocks

- Lack of support from funding bodies
- Lack of incentives
- Publishing data not part of "publish or perish" culture
- Journal articles aren't great for sharing data

Duncan Hull, Why don't scientists share data?

Bryn Nelson, Data sharing: Empty Archives

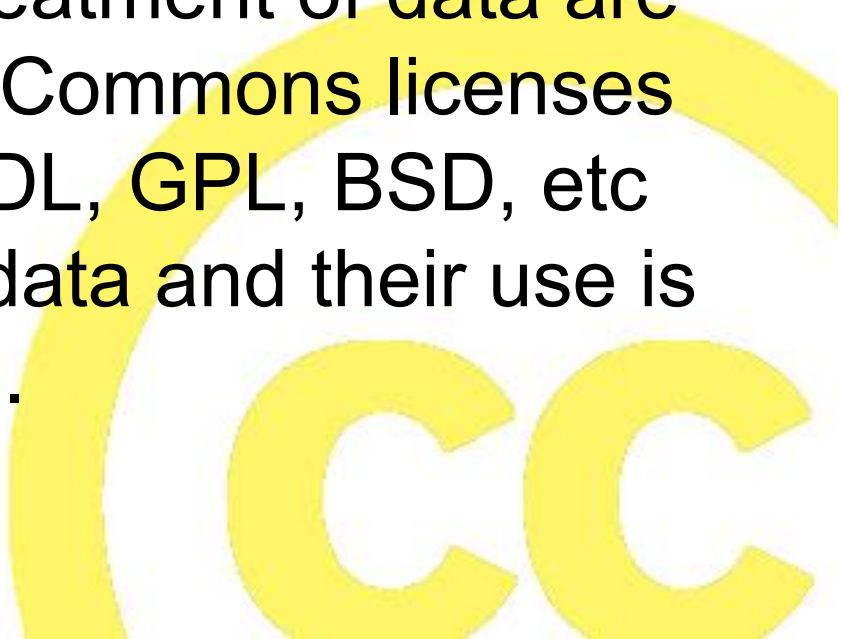
The First Principle

Where data or collections of data are published it is critical that they be published with **a clear and explicit statement of the wishes and expectations of the publishers with respect to re-use and re-purposing of individual data elements, the whole data collection, and subsets of the collection.** This statement should be precise, irrevocable, and based on an appropriate and recognized legal statement in the form of a waiver or license.

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

The Second Principle

Many widely recognized licenses are not intended for, and are not appropriate for, data or collections of data. A variety of waivers and licenses that are designed for and appropriate for the treatment of data are described here. Creative Commons licenses (apart from CCZero), GFDL, GPL, BSD, etc are NOT appropriate for data and their use is STRONGLY discouraged.



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

The Third Principle

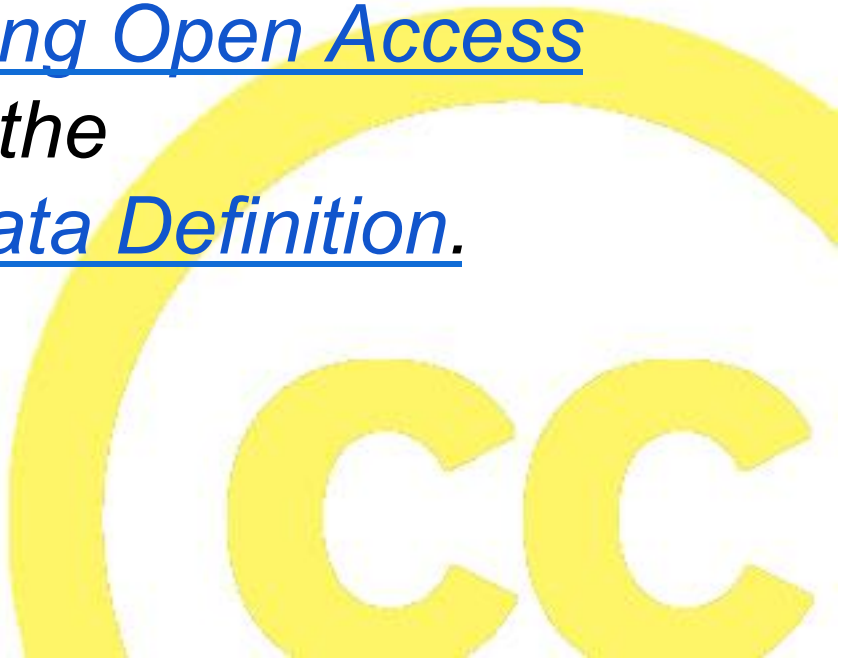
The use of licenses which limit commercial re-use or limit the production of derivative works by excluding use for particular purposes or by specific persons or organizations is STRONGLY discouraged. **These licenses make it impossible to effectively integrate and re-purpose datasets and prevent commercial activities that could be used to support data preservation.**

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.***

The Fourth Principle

Furthermore, in science it is STRONGLY recommended that data, **especially where publicly funded, be explicitly placed in the public domain via the use of the Public Domain Dedication and Licence or Creative Commons Zero Waiver.** This is in keeping with the public funding of much scientific research and the general ethos of sharing and re-use within the scientific community.

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.*



In other words...

1. Make your wishes explicit
2. Use a recognized license or waiver
3. It's better not to use a license or waiver that has restrictive clauses which limit commercial or other reuses
4. In fact, the Public Domain Dedication License or CC0 are the best choices

The Denton Declaration: An Open Data Manifesto

University of North Texas

- Can think of it as almost operationalizing the Panton Principles
- More detailed and and with many more detailed recommendations
- Concentrates both on actions individual researchers can take and on ways research culture should change

UNT is in Denton, TX

- Declarations
  - Managing research data is the responsibility of a broad community of stakeholders including researchers, funders, institutions, libraries, archivists, and the public.
- Principles
  - Open access to research data benefits society, and facilitates decision making for public policy.
- Intentions (ie. advocacy)
  - A culture of openness in research.
- Invitation to sign

An interesting related case study:

Textmining the journal literature

Some links with more info:
http://bit.ly/dupuisdatalinks

Contact Info

John Dupuis
jdupuis@yorku.ca
http://scienceblogs.com/confessions/
http://twitter.com/dupuisj

This presentation: http://bit.ly/DupuisSC13
Sources: http://bit.ly/dupuisdatalinks