sharing scientific data: legal, normative, and social issues

kaitlin thaney
program manager, science commons
beijing, china - 25 march 2009

legal issues:

“it’s complicated”

copyright and ***databases***

what's protected? is it legal?

facts are ***free***

to what extent is there creative expression?

database protections based on jurisdiction

sui generis,
“sweat of the brow”
Crown copyright

the list goes on ....

still murky ...

what is / is not protected?
what rights does the user have?

social issues:

protection instinct / culture of control

PD relinquishes much of this control, even control in the service of freedom

“my data”, interpretation issues

fear, uncertainty, doubt (FUD)

issue of ***license propagation***

whatever you do to the least of the databases, you do to the integrated system

(the most restrictive wins)

need for a ***legally accurate*** and
simple solution

reducing or eliminating the need to make the distinction of what's protected

requires modular, ***standards based*** approach to licensing

our solution -

reconstruction of the ***public domain***

create legal ***zones of certainty*** for data

attribution through ***accompanying norms***

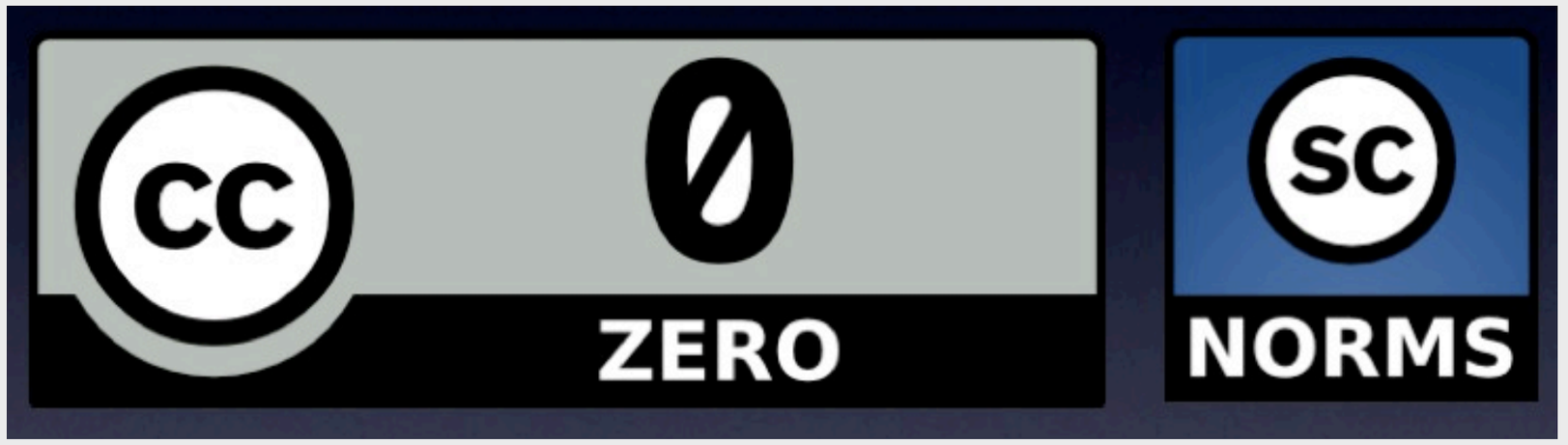
3.1 The protocol must promote legal predictability and certainty.

3.2 The protocol must be easy to use and understand.

3.3 The protocol must impose the lowest possible transaction costs on users.

For the full text:
http://sciencecommons.org/projects/publishing/open-access-data-protocol/

CC Zero waiver + SC norms



waive rights → public domain

+

attribution / citation through community norms, not a contract

science commons

Protocol for Implementing Open Access Data

Status of this Memo

This memo provides information for the Internet community interested in distributing data or databases under an "open access" structure. There are several definitions of "open" and "open access" on the Internet, including the Open Knowledge Definition and the Budapest Declaration on Open Access; the protocol laid out herein is intended to conform to the Open Knowledge Definition and extend the ideas of the Budapest Declaration to data and databases.

This memo does not specify an Internet standard of any kind, but does specify the requirements for gaining and using the Science Commons Open Access Data Mark and metadata, by using legal tools and norms that conform to the protocol specified. This memo is available under the Creative Commons Attribution 3.0 (unported jurisdiction) license and will be submitted to the World Wide Web Consortium for consideration.

The terms MUST, MUST NOT, and SHOULD are used herein as defined in RFC 2119 ("Key words for use in RFCs to Indicate Requirement Levels").

1. Intellectual foundation for the protocol

The motivation behind this memorandum is interoperability of scientific data.

The volume of scientific data, and the interconnectedness of the systems under study, makes integration of data a necessity. For example, life scientists must integrate data from across biology and chemistry to comprehend disease and discover cures, and climate change scientists must integrate data from wildly diverse disciplines to understand our current state and predict the impact of new policies.

The technical challenge of such integration is significant, although emerging technologies appear to be helping. But the forest of terms and conditions around data make integration difficult to legally perform in many cases. One approach might be to develop and recommend a single license: any data with this license can be integrated with any other data under this license.

But this approach, which implicitly builds on intellectual property rights and the ideas of licensing as understood in software and culture, is difficult to scale for scientific uses. There are too many databases under too many terms already, and it is unlikely that any one license or suite of licenses will have the correct mix of terms to gain critical mass and allow massive-scale machine integration of data.

Therefore we instead lay out principles for open access data and a protocol for implementing those principles, and we distribute an Open Access Data Mark and metadata for use on databases and data available under a successful implementation of the protocol.

a protocol, not a license

calls for data providers to ***waive*** all rights necessary for ***data extraction*** and ***re-use***

requires provider place ***no additional*** obligations (like share-alike) to limit downstream use

request behavior (like attribution) through ***norms*** and ***terms of use***

public domain is the natural state of data

examples:
human genome, geographic data,
NASA photographs

public domain ≠ license, cannot be made
“more free” - only less free

PD = the ***original*** commons

at least make metadata open,
if can’t make data itself open

ensure ***freedom to integrate***

kaitlin@creativecommons.org

sciencecommons.org

neurocommons.org