Open Knowledge Foundation
Promoting Open Knowledge in a Digital Age

# Principles on Open Bibliographic Data

**Contributors:** Karen Coyle, Mark MacGillivray, Peter Murray-Rust, Ben O' Steen, Jim Pitman, Adrian Pohl, Rufus Pollock, William Waites

Producers of bibliographic data such as libraries, publishers, universities, scholars or social reference management communities have an important role in supporting the advance of humanity's knowledge. For society to reap the full benefits from bibliographic endeavours, it is imperative that bibliographic data be made open -- that is available for anyone to use and re-use freely for any purpose.

## Bibliographic Data

To define the scope of the principles, in this first part the underlying concept of bibliographic data is explained.

### Core Data

Bibliographic data consists of bibliographic descriptions. A bibliographic description describes a bibliographic resource (article, monograph etc. - whether print or electronic) with the purpose of:

1. *identifying* the described resource, i.e. pointing to a unique resource in the universe of all bibliographic resources and
2. *locating* the described resource, i.e. indicating how/where to find the described resource.

Traditionally one description served both purposes at once by delivering information about: author(s) and editor(s), titles, publisher, publication date and place, identification of parent work (e.g. a journal), page information.

In the web environment identification makes use of Uniform Resource Identifiers (URIs) like a URN, DOI etc. Locating an item is made possible through HTTP-URIs known as Uniform Resource Locators (URLs). All URIs for bibliographic resources thus fall under this narrow concept of bibliographic data.

### Secondary Data

A bibliographic description may include other information that falls under the concept of bibliographic data, such as non-web identifiers (ISBN, LCCN, OCLC etc), rights assertions, administrative data and more[1]; this data may be produced by libraries, publishers, scholars, online communities of book lovers, social reference management systems, and so on.

Furthermore, libraries and related institutions produce controlled vocabularies for the purpose of bibliographic description, such as name and subject authority files, classifications etc., which also fall under the concept of bibliographic data.

[1] See addendum for a list of secondary bibliographic data.

# Four Principles

**Formally, we recommend adopting and acting on the following principles:**

1. Where bibliographic data or collections of bibliographic 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 bibliographic descriptions, 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 bibliographic data make an explicit and robust license statement.***

2. Many widely recognized licenses are not intended for, and are not appropriate for, bibliographic data or collections of bibliographic data. A variety of waivers and licenses that are designed for and appropriate for the treatment of data are described at http://opendefinition.org/licenses/#Data. Creative Commons licenses (apart from CC0), 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.***

3. 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. They furthermore prevent commercial services which add value to bibliographic data or commercial activities which 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 Definition (http://opendefinition.org/) – in particular non-commercial and other restrictive clauses should not be used.***

4. Furthermore, it is recommended that bibliographic data or collections of bibliographic data, especially where publicly funded, should be explicitly placed in the public domain via the use of the Public Domain Dedication and Licence or Creative Commons Zero Waiver. This promotes the maximum possible reuse of the data, in line with the general ethos of sharing within the publicly funded cultural heritage sector.
***Where possible, we recommend explicitly placing bibliographic data in the Public Domain via PDDL or CC0.***

# Addendum

A non-comprehensive list of bibliographic data.

*Core data:* names and identifiers of author(s) and editor(s), titles, publisher information, publication date and place, identification of parent work (e.g. a journal), page information, URIs.

*Secondary data:* format of work, non-web identifiers (ISBN, LCCN, OCLC number etc.), an indication of rights associated with a work, information on sponsorship (e.g. funding), information about carrier type, extent and size information, administrative data (last modified etc.), relevant links (to wikipedia, google books, amazon etc.), table of contents, links to digitized parts of a work (tables of content, registers, bibliographies etc.), addresses and other contact details about the author(s), cover images, abstracts, reviews, summaries, subject headings, assigned keywords, classification notation, user-generated tags, exemplar data (number of holdings, call number), ...