

WorldCat Data Licensing

1. Why is OCLC recommending an open data license for its members?

Many libraries are now examining ways that they can make their bibliographic records available for free on the Internet so that they can be reused and more fully integrated into the broader Web environment. Libraries may want to release catalog data as [linked data](#), as MARC21 or as MARCXML. For an OCLC® member institution, these records may often contain data derived from WorldCat®. Coupled with a reference to the community norms articulated in [WorldCat Rights and Responsibilities for the OCLC Cooperative](#), the Open Data Commons Attribution ([ODC-BY](#)) license provides a good way to share records that is consistent with the cooperative nature of OCLC cataloging.

Best practices in the Web environment include making data available along with a license that clearly sets out the terms under which the data is being made available. Without such a license, users can never be sure of their rights to use the data, which can impede innovation.

The [VIAF](#) project and the [addition](#) of Schema.org linked data to WorldCat.org records were both made available under the ODC-BY license. After much research and discussion, it was clear that ODC-BY was the best choice of license for many OCLC data services. The recommendation for members to also adopt this clear and consistent approach to the open licensing of shared data, derived from WorldCat, flowed from this experience.

An OCLC staff group, aided by an external open-data licensing expert, conducted a structured investigation of available licensing alternatives to provide OCLC member institutions with guidance. Before this recommendation was adopted, OCLC Global Council considered the conclusions of the OCLC staff group and approved this direction, as did the OCLC Board of Trustees.

2. How does the ODC-BY license relate to WorldCat rights and responsibilities?

WorldCat Rights and Responsibilities for the OCLC Cooperative (WCRR) was the result of a consultative process led by OCLC member institutions and ratified by our Global Council and the OCLC Board of Trustees. The WCRR reflects the understanding of the stakeholders in WorldCat on how they'd like to see WorldCat sustained.

In line with WCRR, OCLC's current recommendation to OCLC members thinking about releasing their catalogs containing WorldCat-derived records is that they consider using an ODC-BY license. The ODC-BY license requires attribution of OCLC, WorldCat and the member institution, should they require it. OCLC recommends that members use ODC-BY coupled with a statement that the use and transfer of the WorldCat-derived records in the released database (i.e., the catalog) should comply with WCRR.

Use of an ODC-BY license coupled with a statement that use of WorldCat-derived records in the released catalog as described above is consistent with the obligations set forth in WCRR.

3. Why is OCLC asking for attribution for WorldCat records?

OCLC members believe that attribution of WorldCat and OCLC helps keep WorldCat as a sustainable resource for both OCLC members and the general public in line with WCRR.

WorldCat represents a significant investment of time and resources both from OCLC and from each of its member institutions that contribute records. The WorldCat community depends on members to demonstrate the value of WorldCat to the institutions in the OCLC cooperative by attributing WorldCat as the source. By attributing WorldCat or citing WorldCat in applications, websites, news stories, articles and research reports, members can directly impact the cooperative's ability to continue providing data to the community.

4. Why is an open license for WorldCat data recommended?

Open-source software licensing and open-content licenses have made great strides in developing stable, useful software and rich content available for the general public. The goal of open data is to bring this established approach of open licensing to data.

A key feature of these approaches is that the work should be available for further use by the public without the need to seek further permission from the rights holder. The rights holder up front gives the world a license to use their work. This is important as negotiating individual licenses can waste resources (for both the rights holder and the user), especially when the rights holder would like to share their information but has no easy way of doing so.

5. Why is there “the database versus contents” distinction?

Often when people talk about data they mean only factual information—water boils at 100 Celsius, for example. But in the context of data licensing, the contents of a database can be anything, including

- mobile video;
- images, such as photos from Flickr;
- text documents, such as Microsoft Word .doc files; and
- factual information, such as dates or budget numbers.

Anything that can be stored in a database can be “data” in this context. It can be helpful to distinguish between the “contents of a database” and “the database.” This is the data/database distinction, and from a legal perspective it means that there may be legal rights specific to the contents of a database (the data, because it can be anything). Also, any collection of contents (the database and potentially even an XML file) may have legal rights covering that collection that are independent of the rights over individual contents within the collection.

For example, if you had a database of still images and video of science fiction movies from the 1960s and 1970s, you may have rights, including for your selection and arrangement of the sci-fi films, over the database as a whole. And you would still have to clear the rights (such as copyright) over the images and video you included inside your database.

One rights layer might exist over the contents (records, in the library-use case) and another rights layer might exist over the database (the compilation, field names, schema and other aspects of the database apart from the records).

This is the reason why more than one license might be necessary to cover both the database and the contents.

6. Can “attribution stacking” be a problem with requiring attribution?

Often, the question will come up when recommending an attribution license as to whether a public domain approach is preferable due to the “attribution stacking problem.” The argument is that since database contributions can come from a large number of people, giving credit doesn’t scale. For example, one may have a database of a million records from a million different sources. The thought is that implementing attribution requirements could be difficult and expensive (in terms of file size, bandwidth, time to manage, etc.). Public domain approaches are often put forward as a solution, as they would not require attribution at all (nor any other license compliance terms).

It’s important to note that attribution stacking problems don’t go away, even with public domain for a database, if the contents are all under different licenses (Flickr photos, music, etc.).

In the case of attribution of a database, any accumulation of attributions will occur outside the individual information objects within the database. A service or database built from many sources that require attribution might need to have a separate webpage listing the sources and the required form of attribution that can be pointed to by others.

7. Do I have to remove OCLC control numbers (OCNs) when sharing bibliographic records with partners?

No. It is in the interest of libraries and the cooperative for OCLC control numbers to be widely disseminated. Therefore, you don’t need to remove OCNs when exposing or exchanging bibliographic data with partners.

8. If we can treat OCNs as if they are in the public domain, why doesn’t OCLC explicitly make them public domain?

The OCN is an individual bibliographic element, and OCLC doesn’t make any copyright claims either way on specific data elements. The OCN can be used by other institutions in ways that, at an aggregate level, may have varying copyright assertions. Making a positive, specific claim that the OCN is in the public domain might interfere with the copyrights of others in those situations.