|  |  |
| --- | --- |
| **Why, when, and how to cite sources of biodiversity specimen data**Pearson KD. 2020. |  |

Link to video: <https://youtu.be/z7E6VvtNvpY>

**Video Script**

Natural history collections data, that is, images of and data from preserved plants, animals, and other once-living creatures are becoming more available than ever with the rise of digitization efforts. Institutions across the globe that house scientific specimens are imaging, cataloging, transcribing, and geo-referencing millions of specimens, and these millions of data points are empowering new heights of scientific research, education, and general availability of information to the public.

Many of these data sources are free and open to the public. Anyone with internet access can search huge databases and download thousands of records at a time. However, just because these data are free to use doesn't mean they are free from responsibility. It is vital to accurately cite the source of any natural history collections data that you use.

But why is this the case? We'll discuss four major reasons in this video: legal reasons, scientific integrity and reproducibility, giving proper credit for other people's work, and increasing the visibility of the data to help justify the costs of creating and maintaining the collections that made those data available in the first place.

First, using data without proper attribution is usually against copyright law depending on the data license. Even Creative Commons licenses usually include an attribution element, meaning that legal use of the data requires the user to give attribution to (aka to cite) the data source in any products that use the data.

Beyond legal reasons, citing your data sources is part of basic scientific ethics. One key reason is that it enables your scientific research to be reproducible. Any other scientist could discover the same results that you did if they gathered the same data and did the same analyses. This is important for scientific integrity transparency and quality. If you don't cite your data source, who's to say you didn't just make it all up so that it says what you want it to say? That's not science.

The reality is, these data are the hard-won result of dedicated collectors, curators, directors, digitizers, volunteers, students, and others, and citing your data source provides critical recognition of their hard work.

Not only does citation increase the visibility of a collection or other data source, but it also helps collections justify the costs of maintaining and digitizing their specimens. Getting credit for data use helps collections attract potential funders and therefore create more and better data, or in some unfortunate cases, just to justify their existence in the face of potential shutdowns or budget cuts prolonging the life and utility of the data.

Okay, so we've made it clear that citing data sources is important, but when should you do it, and how should you do it? You should cite the source of your data whenever you are presenting the data or analyses that use the data, that means presentations, posters, scientific papers, and class assignments. All of these products should have a references or works cited section in which you cite the data source.

You should also include an in-text or parenthetical citation when you refer to the data source. For example, if you analyze natural history specimen data for your research project, you should cite the data source in the text of your methods section. If you are just using natural history specimen data as background information for a presentation or other project, you should also cite the data where you refer to it. If you use an image of a natural history specimen collection, you should provide a brief citation near or around the image what should the citation look like well it depends on the type and source of the data. For images of natural history specimens, pay attention to the image license. Most often, the image will be protected under a Creative Commons license. Make sure to cite the image as directed by the specific license. For example, the images we showed earlier of natural history collections are protected by a CC-BY license. This means that we need to provide the name of the contributor and a link to the license applied to that image. You can find more information about how to cite images protected by Creative Commons licenses on the Creative Commons website.

The proper format for citing other natural history specimen data depends on the data source. To find out how a data source prefers to be cited, take a look at the source's data use policy, terms of use, or how to cite page. For example, data downloads from GBIF automatically produce a recommended citation with a unique URL to the search that you conducted. For Symbiota-based portals, such as SEINet, CCH2, SERNEC, and many others, the recommended citation is usually listed on the data use policy page. It's as easy as copying and filling out a template.

Other collections data portals may or may not have easy-to-find citation recommendations, but that doesn't mean you aren't responsible for citing the data. Try entering "how to cite" and the name of the data source into a search engine, or if all else fails, you can email the data sources curator or webmaster for guidance. Chances are, they will be happy to point you to the right place.

We as a community of data providers, data curators, and data users all have responsibilities to each other to give appropriate credit to the people and institutions that make data discovery and analysis possible. After all , with any luck someday someone may cite you!