Data Licensing for the NOAA Office of Coast Survey

Matt Wilson¹, James Miller¹, Glen Rice¹, Kim Valentine², Meredith Westington¹, Katrina Wyllie¹

¹ NOAA Office of Coast Survey, U.S.A.
² NOAA National Ocean Service, U.S.A.

matthew.wilson@noaa.gov

Abstract

The NOAA Office of Coast Survey (OCS), working in collaboration with the Office of General Counsel and the Information Management Office, is exploring data license options for receiving bathymetry from the ocean mapping community.

While the data that OCS acquires directly (or has contracted to acquire) are usually in the public domain with no restrictions on use, a major part of the NOAA mapping strategy involves taking advantage of External Source Data (ESD), which is bathymetry acquired by separate entities that has value to OCS even though it may not have been intended for nautical charting or related applications. These providers are often willing to contribute their data to NOAA at no cost, but may have nuanced data restrictions or existing licenses which impact data handling, application to OCS products, and public access.

Informed by the NOAA Data Strategy and guidance from the International Hydrographic Organization, OCS intends to develop a mechanism that will enable providers to apply well-known, universal licenses from the Creative Commons and Open Data Commons to the data prior to submission. The licenses would define the permitted uses and applications of the data and derived products. Certain products such as Electronic Navigational Charts (ENCs) are not capable of carrying data licenses and therefore only include open data with no restrictions on use. However, other products such as the National Bathymetric Source (NBS), the source database for NOAA nautical charts, was built to store and convey the data provider information and the license attributed to each dataset. The benefits of adopting the data licenses could include improved communication to users and standardized handling of the data that will increase efficiency, while minimizing potential errors and effort.

Background

The data acquired by NOAA hydrographic survey units and contractors are simply not enough to achieve our various national and global mapping objectives. We must rely on External Source Data (ESD), which OCS has defined as data opportunistically accessed from public repositories or voluntarily submitted to NOAA. OCS increasingly relies on ESD to supplement our mapping efforts, and this is stated explicitly in the current OCS Strategic Plan (2020) under objective 2.1.3, which is to “increase the speed, quality, consistency, and transparency of external source data ingest to Coast Survey.” The importance of ESD is
further underscored in the National Strategy for Mapping, Exploring, and Characterizing the United States Exclusive Economic Zone (NOMEC, 2020) and the Seabed 2030: A Call to Action (Smith, 2018), which both emphasize the critical importance of collaboration and leveraging data from external sources in order to achieve our mapping goals.

Depending on the quality of the data, ESD is used for a variety of applications within OCS such as assessing chart adequacy, survey reconnaissance, or application to navigational products. The goal is to use existing data and coordinate survey plans to minimize redundancy between NOAA, our federal partners, and the industry at large, and thereby increase our efficiency and save considerable time and resources. When permitted by the data provider and applicable law, data shared with NOAA are publicly archived and made openly available to support national and global mapping strategies, fill bathymetric gaps, and make the data discoverable for various other applications.

The Challenges of External Source Data

The application of external data to OCS products and the ensuing archive of the data for public consumption as discussed above is entirely dependent upon the data being free from any restrictions. Without a standardized data licensing framework, data restrictions are disruptive to the efficient use and public distribution of the data and derivative products. Such restrictions also represent a fundamental incongruence with U.S. government data, which is federally mandated to be accessible and open (44 U.S.C. § 3506(b)(6)); thus ESD with restrictions requires an entirely different pathway, consuming substantial time and resources to ensure its proper handling.

ESD is often received with restrictions on the use and access of data. Such restrictions may be defined in a data sharing agreement, commonly referred to as a Memorandum of Understanding or Agreement (MOU or MOA). While these agreements are a useful way to facilitate data sharing, it may be beneficial to instead use data licenses, as explained further below.

When is Intellectual Property Subject to Copyright?

Prior to discussing the licensing of data, it is pertinent to acknowledge intellectual property, when it may be subject to copyright, and how it applies to data. Black’s Law Dictionary (2019) defines intellectual property as a category of intangible rights protecting commercially valuable products of the human intellect. There are four ways to protect intellectual property (USPTO, 2021), including copyright, which allows the owner the right to say how others may use their intellectual property. In the U.S., copyrights are approved by the U.S. Copyright Office at the Library of Congress. U.S. copyright law protects authors of original works, including maps. As noted on NOAA’s Electronic Navigational Charts (ENCs), no copyright is claimed by the United States Government under Title 17 U.S.C. (U.S. Copyright Office, 2021). However, other nations may claim intellectual property rights on the compilation of data depicting waters not subject to U.S. jurisdiction.

How copyright law applies to data comprising measurements of the seafloor is less clear, though the Evidence Act (44 U.S.C. § 3506(b)(6)) would support an understanding that all government data, like products, are to be made public, except when the data is exempt from public release under the Freedom of
Information Act (5 U.S.C. § 552) or other applicable laws. Non-federal data collectors, such as industry, possess many data that are subject to restrictions in order to protect business confidential information. Similarly, academia also collects data (often in partnership with the federal government) and may restrict its access for a number of years during the ensuing research. Because the link between data and intellectual property rights can be unclear, the application of a data license may help to remove ambiguity and add clarity to how data may be used (Open Knowledge Foundation, 2021).

Why Data Licensing?

Developing data licensing guidance is listed as part of objective 3 in the NOAA Data Strategy (2020). The associated milestones and deliverables (Collins, Valentine, and Hanson, 2021) specify licensing guidance to be machine-readable and to clearly define characteristics that conform with and support the Evidence Act requirements (44 U.S.C. § 3506(b)(6)) as they pertain to U.S. government data, and to develop a NOAA policy for specifying appropriate data licenses for data that NOAA might receive from external sources (i.e., ESD). Taken further, the U.S. government addresses the use of licenses towards proper stewardship of internal and external data (resources.data.gov, 2021 and OMB M-13-13, 2013). Even though U.S. government data that may theoretically be considered intellectual property are not subject to copyright per Title 17 U.S.C. (U.S. Copyright Office, 2021), that open policy does not necessarily apply to data users outside of the U.S. Thus, while open data licenses are not strictly necessary within the U.S., their application could help reduce any confusion about the use of U.S. government data outside of the U.S. This is particularly relevant given that U.S. bathymetry data and products are available worldwide, thus adhering to international standards in the distribution and terms of use is deemed beneficial.

There are clear benefits to the data licensing approach. Data received under a standard license would have the associated uses and access of the data clearly defined by the license terms. This approach would effectively standardize data handling and facilitate automation, resulting in faster propagation of ESD through the pipeline while minimizing (or eliminating altogether) any potential errors or mishandling. Additionally, because the licenses would be modeled after well-known, universal license types that users already are (or will become) familiar with, there would be improved communication to users upon recognition of the license and understanding of the allowable applications of a particular dataset. There could be further benefits from the standardization when datasets are combined, as the data licenses facilitate legal interoperability, meaning that the resulting dataset will incorporate the accumulated restrictions imposed by each source (RDA-CODATA Legal Interoperability Interest Group, 2016).

Why Coast Survey?

As discussed above, data licensing is a priority for the NOAA Data Strategy (2020). OCS is incentivized to explore data licensing options, considering the challenges associated with ESD ingest that are inhibiting our greater mapping objectives. Thus, OCS is working closely with the NOAA National Ocean Service’s Information Management Office and the NOAA Office of General Counsel to ensure that data licenses would be legally and structurally sound.
In anticipation of the adoption of data licensing, OCS has been developing the infrastructure to best accommodate the licenses to fully realize the associated benefits. Specifically, the National Bathymetric Source (NBS), Coast Survey’s source database for the best available bathymetry (Wyllie and Rice, 2020), was built with the infrastructure to hold and maintain data licenses as part of a georeferenced attribute table, as shown in Figure 1.

![Figure 1. The National Bathymetric Source was from its inception built to maintain source attribution and licensing. The seamless layer of bathymetry (left) is composed of numerous sources (right), queried in the table below to show the corresponding licensing and source identification.](image)

The primary data contained within the NBS is the open-source Bathymetric Attributed Grid (BAG), which contains georeferenced elevation and uncertainty, but only general metadata applicable to the entire dataset (i.e., not georeferenced). To better support the data licensing paradigm (and S-101 products), OCS issued a proposal to the Open Navigation Surface Working Group (ONSWG), the group that supports the BAG, to add a georeferenced metadata layer to the BAG that specifies the quality of bathymetry and data licensing (Rice, 2019). An implementation of the proposed BAG specification is underway by Teledyne CARIS, with funding from NOAA. With the NBS carrying our primary data format (BAG), updated to contain the georeferenced metadata layer describing the licensing, OCS will be well-positioned to best realize the benefits of data licensing.

**License Adoption**

While the benefits and motivations for data licensing are clear, the question now is what licenses should OCS adopt? Ideally, the data license options would directly align with the various data uses and products that OCS creates, so that license options would cover the full spectrum of restriction scenarios that are
encountered with external data providers. Such an ideal suite of license options is only possible if custom-made licenses are created by OCS and tailored for such scenarios. However, this approach could be challenging since custom data licenses would not be broadly applicable to the industry or to other NOAA offices that wish to adopt data licensing.

Rather, the approach OCS is most interested in pursuing is the well-known and established licenses from Creative Commons and Open Data Commons. The licenses are internationally recognized and already widely used across a variety of industries. Though the “fit” of the individual license types is not as clean as it would be for custom-made licenses, the broad recognition of the licenses by the industry at-large would likely be deemed more valuable. Providing more incentive yet to the adoption of the Creative Commons licenses is the guidance from the International Hydrographic Organization (2018) that not only describes why data needs to be licensed, but mentions explicitly the licenses from Creative Commons and Open Data Commons as standard licenses for governments and organizations to adopt. Finally, several notable bathymetric databases in the hydrographic community have already applied Creative Commons licenses. For example, the Global Multi-Resolution Topography (GMRT, 2021) synthesis from the Lamont-Doherty Earth Observatory of Columbia University and several datasets available through the Land Information New Zealand Data Service (LINZ, 2021) are each licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Given the acknowledgement from the International Hydrographic Organization, whom hydrographic offices worldwide and the industry at-large commonly look to for guidance, and considering the existing familiarity within the industry, licenses from Creative Commons and Open Data Commons seem to be the preferred approach.

Open Data Licenses

As mentioned above, U.S. government data is mandated to be open and accessible (44 U.S.C. § 3506(b)(6)). In terms of licensing such data, the Federal Open Data Policy states “agencies must apply open licenses, in consultation with the best practices found in Project Open Data, to information as it is collected or created so that if data are made public there are no restrictions on copying, publishing, distributing, transmitting, adapting, or otherwise using the information for non-commercial or for commercial purposes” (OMB M-13-13, 2013). The Creative Commons 1.0 Universal Public Domain Dedication (CC0) and the Open Data Commons Public Domain Dedication and License (PDDL) are both listed on resources.data.gov (2021) as examples of open licenses and thus may be acceptable for data collected by OCS, either directly by NOAA hydrographic field units or under the hydrographic services contract.

Derivative Products

In the workflow of a particular organization, it is critical to understand what is or is not a derivative product, because this will impact if the creation and/or public distribution of the product is permissible under a particular license from Creative Commons. Additionally, it is critical to understand if the derivative product can maintain the same license assigned to the original data.
From the original licensed source data, products will be created. If the data is fundamentally changed in order to make the product, or if it cannot be easily reverse-engineered, then the product is considered to be derivative. An example in the context of nautical charting is when a BAG is used as an input to create the ENC. In this case, the data is fundamentally changed (having gone from a raster to a vector format), thus it is not easily reverse-engineered and, therefore, the ENC is considered a derivative product of the original data. Additionally, the ENC is a public domain product and has no mechanism with which to maintain the data license assigned to the original data. Thus, data provided under a license with “no derivatives” or “sharealike” restrictions could not be used to create an ENC (see below).

License Restrictions

The Creative Commons licenses (Creative Commons, 2021) have four key restrictions that define the various license types. With the exception of the CC0 Public Domain Dedication, each of the licenses from Creative Commons has one or more of these restrictions. They are defined below, and also given is the abbreviation used to denote the restriction.

- **Attribution (BY):** credit must be given to the provider for the original data.
- **Sharealike (SA):** any derivative products created from the original data must carry the same license.
- **No Derivatives (ND):** derivative products may not be created from the original data.
- **Non-commercial use only (NC):** the original data may only be used for non-commercial purposes.

Coast Survey uses for ESD

Below are the prominent uses for the ESD that OCS receives. This is not an exhaustive list, but rather they represent the primary categories of uses that OCS has for ESD that would be affected by the various Creative Commons license types.

- **Internal use only:** the data is kept internally within NOAA (not just OCS), and is used for planning purposes, internal modeling, and other applications that are not external-facing.
- **Public archive:** non-sensitive data is posted in its original form for public access, most commonly at the National Centers for Environmental Information (NCEI).
- **National Bathymetric Source (NBS):** non-sensitive data is merged into the OCS source database for best available bathymetry. Note that, as shown in Figure 1, the NBS holds the data in its original form and maintains both the source attribution and data license; thus, the NBS is not considered a derivative product, and would be permitted under sharealike.
- **Modeling, ENC, Precision Navigation, etc.:** this is a catch-all for the wide variety of products created by NOAA that are all considered derivative and that are in the public domain, thus they would not be permissible under the “sharealike” and “no derivatives” restrictions.
License Options

Listed below are the licenses that OCS is considering making available to providers to define how OCS may use it. For full information, including the license deeds, see the Creative Commons and Open Data Commons websites. If needed, OCS could likely use a custom-made license for data submissions that would be for internal use only, in the case that the data are exempt from public release.

- **Creative Commons 1.0 Universal Public Domain Dedication (CC0)**: the data is open and can be freely used, modified, and shared by anyone and for any purpose.
- **Open Data Commons Public Domain Dedication and License (PDDL) v1.0**: offered as an alternative to CC0, for convenience in case that a particular provider has already assigned their data as PDDL.
- **CC BY**: a license requiring attribution, meaning credit must be given to the provider for the original data, but otherwise it can be freely used, modified, and shared by anyone for any purpose. Derivative products may be created and are in the public domain.
- **CC BY-SA**: like CC BY, but requires sharealike. The original data can be freely used and shared by anyone for any purpose, but derivative products are prohibited.
- **CC BY-ND**: like CC BY, but specifies no derivatives. The original data can be freely used and shared by anyone for any purpose, but derivative products are prohibited.
- **CC BY-NC**: like CC BY, but requires non-commercial use only, meaning that the data can be freely used and shared, but for non-commercial purposes only. Derivative products may be created and are in the public domain.
- **CC BY-NC-SA**: like CC BY-NC, however, the sharealike prohibits any derivative products.
- **CC BY-NC-ND**: like CC BY-NC, however, no derivatives prohibits any derivative products.
- **Internal Use Only**: the data will not be shared, and we will not create derivative products from it. The data will be for internal use only.

For ESD providers, OCS would encourage open data submissions (e.g. CC0 or PDDL) as the fastest and most flexible type of data submission.

Reference Table 1 for OCS uses of data matched against the various potential license options, where it is shown which uses are likely authorized and which are likely prohibited under each license type.
Sensitve Data Handling

Over the years, OCS has periodically received ESD that is sensitive in nature, and exempt from public release under applicable laws such as the Freedom of Information Act (5 U.S.C § 552) and National Historic Preservation Act (54 U.S.C. §§ 300101 et seq.). Generally, the original source data would be labeled as sensitive and placed in a non-public archive, and if it was permissible to use the data for derivative products, generalized versions of the data or its cartographic components (e.g., soundings and contours) would be extracted for chart compilation, while the original source data remained in a non-public archive. Any sensitive features would not be publicly shared, although least depths would be retained in the sounding selections to ensure safety of navigation.

In the data licensing paradigm, the process would be similar. Sensitive datasets would receive an Internal Use Only license and be sent to the non-public archive. In the case of sensitive features, OCS has a sensitive data policy to allow for “redaction” of those particular features, which allows for the redacted dataset to enter the public domain with the sensitive features effectively masked. In this case, there are effectively two versions of the dataset: the original dataset is assigned an Internal Use Only license, whereas the redacted version is assigned a less restrictive license (preferably an open data license).

Summary and Next Steps

Currently, OCS is working on a draft policy and procedures to inform and standardize future use of data licenses. OCS also continues to communicate with internal and external stakeholders and partners, and at various conferences, to solicit feedback on the potential use of data licenses.

The work done thus far has focused on two simple questions: in our workflow, what is and what is not considered a derivative product, and can that product hold a data license? With clear answers to those questions, the application of the well-known and universal license types discussed here would be fairly straightforward. The more significant work is developing the capability for databases, data formats, and products to hold data licenses in order to gain the most benefit from this approach.
References


