RECOMMENDED BEST PRACTICES FOR BETTER SHARING OF CLIMATE DATA

by Taylor Campbell, Wanying Li, and Dr. Cable Green

January 2024

This work is licensed CC BY 4.0.

Suggested attribution:

“Recommended Best Practices for Better Sharing of Climate Data” by Taylor Campbell, Wanying Li, and Dr. Cable Green for Creative Commons is licensed under CC BY 4.0.”

creative commons
OPEN CLIMATE DATA
Facilitating Better Sharing Worldwide

TABLE OF CONTENTS (Click to jump to that section)

PREVIEW THE RECOMMENDATIONS

INTRODUCTION

PART 1: LEGAL AND LICENSING TERMS

- TWO RECOMMENDED OPTIONS
- THE REALITIES OF LICENSE STACKING AND ATTRIBUTION STACKING

PART 2: METADATA VALUES FOR SHARING, ATTRIBUTION, AND PROVENANCE

PART 3: GENERAL MANAGEMENT AND GOVERNANCE ACTIONS

APPENDIX

- DEFINITIONS
  - Copyright Terms
  - Role Types
  - Data Types
- COPYRIGHT BY JURISDICTION
- REFERENCES

SUMMARY

PREVIEW THE RECOMMENDATIONS

RECOMMENDED LEGAL AND LICENSING TERMS FOR YOUR CLIMATE DATA

Option A:

CC0 + ATTRIBUTION REQUEST

"This work is dedicated to the public domain, marked by CC0 1.0. Attribution is requested."

Recommended for:
- Organizations developing new data policies
- Those willing to share their data freely to maximally reduce friction in use

What it is:
- CC0 is not a license; it dedicates climate data to the global public domain
- We add attribution requests by default, although they are not legally required

Benefits:
- Enables distribution, remixing, and adaptation without legal restrictions
- Encourages professional attribution
- Eliminates legal complexities, releasing any burden on users

Considerations:
- Address sensitive data proactively before sharing in the public domain
- Manage the use of name trademarks if necessary

Option B:

CC BY 4.0

"This work is licensed under CC BY 4.0."

Recommended for:
- Producers and publishers that want to share and still own their copyrightable data, where allowed

What it is:
- CC BY is a highly permissive license that allows unrestricted reuse as long as proper attribution is given

Benefits:
- Retains data ownership in permitting jurisdictions
- Offers a legal mechanism for enforcing attribution and proper data use
- Could provide a way to track data use via attributions

Considerations:
- Sharing can become complex when users want to merge with other licensed data
- May not apply in all jurisdictions or on non-copyrightable material

creative commons
OPEN CLIMATE DATA
Facilitating Better Sharing Worldwide

RECOMMENDED METADATA VALUES FOR THE SHARING, ATTRIBUTION, AND PROVENANCE OF YOUR CLIMATE DATA (PAGE 1: ABOUT YOUR DATA)

Group	Metadata Value Label	Example Value
ABOUT MY DATA/ DATABASE/ PRODUCT *My organization has published something and made it available to share with you, and we want to make sure you attribute it to us and our sources.*	Title	Korean Land and Water Levels 2003
	Publisher	Korea Meteorological Administration
	Identifier	doi:XXXXXXX/kma.domain
	License	http://creativecommons.org/publicdomain/zero/1.0
	Rights Statement	CC0 + Attribution Request. This work is dedicated to the public domain, marked by CC0, which allows users worldwide to distribute, remix, adapt, and build upon the material in any medium or format, with no legal conditions on reuse. Attribution is requested in a standard format to allow reusers to include this information easily.
	Bibliographical Citation	Korea Meteorological Administration, Korean Land and Water Levels 2003, Issued 2004-01-01, Modified 2004-03-01, doi:XXXXXXX/kma.domain, http://creativecommons.org/publicdomain/zero/1.0, CC0 + Attribution Request

creative commons
OPEN CLIMATE DATA
Facilitating Better Sharing Worldwide

RECOMMENDED METADATA VALUES FOR THE SHARING, ATTRIBUTION, AND PROVENANCE OF YOUR CLIMATE DATA (PAGE 2: ABOUT YOUR SOURCES)

Group	Metadata Value Label	Example Value
ABOUT MY SOURCE(S) *My organization got data from other sources, and we want to make sure we attribute our sources.*	Source 1 Title	KMA Korean Land and Water Levels 2003
	Source 1 Publisher	Climate Data Store (CDS)
	Source 1 Identifier	doi:XXXXXXX/cds.domain
	Source 1 License	http://creativecommons.org/publicdomain/zero/1.0
	Source 1 Bibliographical Citation	KMA Korean Land and Water Levels 2003, Climate Data Store, doi:XXXXXX/cds.domain, http://creativecommons.org/publicdomain/zero/1.0, CC0 + Attribution Request; Korean Land and Water Levels 2003, Korea Meteorological Administration, doi:XXXXXXX/kma.domain, http://creativecommons.org/publicdomain/zero/1.0, CC0 + Attribution Request
	Source 2 Title	Japanese Land and Water Levels 2003
	Source 2 Publisher	Japan Meteorological Agency
	Source 2 Identifier	doi:XXXXXXX/jma.domain
	Source 2 License	https://creativecommons.org/licenses/by/4.0/
	Source 2 Bibliographical Citation	Japanese Land and Water Levels 2003, Japan Meteorological Agency, doi:XXXXXXX/jma.domain, https://creativecommons.org/licenses/by/4.0/, CC BY 4.0.

creative commons
OPEN CLIMATE DATA
Facilitating Better Sharing Worldwide

FULL REPORT

INTRODUCTION

Climate change and the resulting harm to our global biodiversity is one of the world's most pressing challenges. To address the climate crisis, **we must open and share climate data** that is produced and published, so that it is simple to access and use.

Climate Data Producers and Publishers/Hosts

Inter/Governmental Agencies and Organizations	**Sponsored Research Institutions**	**Community Groups and Individual Researchers**

These recommendations for better sharing of climate data have been drafted by Creative Commons to provide guidance that maximizes access, sharing, and reuse of open climate data, with practical steps that can be implemented today. We developed these recommendations in consultation and collaboration with key members of **government agencies and intergovernmental organizations** that have taken a strong interest in open practices to produce and publish climate datasets (including observation records), data catalogs, and/or data products (including interpretations and analyses), including:

- European Centre for Medium-Range Weather Forecasts (ECMWF)
- International Organization for Standardization (ISO)
- U.S. National Aeronautics and Space Administration (NASA)
- U.S. National Oceanic and Atmospheric Administration (NOAA)
- World Meteorological Organization (WMO)
- World Resources Institute (WRI)

We would like to thank the individuals that consulted with us and contributed to our knowledge for these recommendations, and assert that their participation is solely that and does not inherently represent a commitment to organizational endorsement.

These recommendations are inspired by and intended to provide further guidance to implement existing policies and principles, including the FAIR Principles and Assessment, CARE principles, GEO Data Sharing Principles,[1] and the WMO Unified Data Policy.[2] Our recommendations cover open data, which are typically freely offered by publishers without restrictions. Separately, and under similar principles, the WMO and other data policies often refer to what they designate as "Essential Data" or "Core Data". However, data with more specific restrictions might fall outside the scope of our recommendations (e.g. “Recommended” data in the WMO Unified Data Policy).

In early 2023 we conducted a comprehensive landscape analysis (blog post) of major climate data sources, assessing their current status and examining diverse data-sharing approaches from entities such across governments, global NGOs and intergovernmental alliances. Our goal was to understand and facilitate more effective sharing of climate data in accordance with FAIR principles[3]: findability, accessibility, interoperability and reusability. Anyone interested in sharing climate data can join the many climate data providers and producers around the world who are making their data, databases, data products, metadata, and infrastructure available in ways related to each of these variables.

Findability	F1. (Meta)data are assigned a globally unique and persistent identifier, like DOIs or a standard PID
	F2. Data are described with rich metadata (defined by R1 below)
	F3. Metadata clearly and explicitly includes the identifier of the data they describe
	F4. (Meta)data are registered or indexed in a searchable resource (e.g., federated searches)

[1] GEO Data-Sharing Principles state that data, metadata and products will be shared as Open Data by default, as part of the GEOSS Data Collection of Open Resources for Everyone (Data-CORE) without charge or restrictions on reuse, subject to the conditions of registration and attribution when the data are reused.

[2] WMO Unified Data Policy states that members shall provide on a free and unrestricted basis the core data that are necessary for the provision of services in support of the protection of life and property and for the well-being of all nations, which are required to monitor and predict seamlessly and accurately weather, climate, water and related environmental conditions.

[3] FAIR data isn’t inherently required to be open. In our assessment, we are going beyond FAIR, to also encourage climate data to be open. See: https://www.go-fair.org/resources/faq/ask-question-difference-fair-data-open-data/

	(Also assessed by CC) Database and/or infrastructure has its own search function
Accessibility	A1. (Meta)data are retrievable by their identifier using a standardized communications protocol
	A1.1 The protocol is open, free, and universally implementable
	A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
	A2. Metadata are accessible, even when the data are no longer available
	(Also assessed by CC) Data is available to users free of charge
	(Also assessed by CC) No registration is required in order to access the data
Interoperability	I1. (Meta)data uses a formal, accessible, shared, and broadly applicable language for knowledge representation, in the form of machine-readable file types
	I2. (Meta)data use vocabularies that follow FAIR principles
	I3. (Meta)data include qualified references to other (meta)data
	(Also assessed by CC) Data is downloadable wherever possible
	(Also assessed by CC) No special software is required to access the data
	(Also assessed by CC) All data is hosted locally and does not require access to a separate/third party host with less access.
Reusability	R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
	R1.1. (Meta)data are released with a specific, clear, and accessible data usage license
	R1.2. (Meta)data are associated with detailed provenance
	R1.3. (Meta)data meet domain-relevant community standards

	(Also assessed by CC) Base data is either dedicated to the public domain using CC0 with Attribution Request or is openly licensed with CC BY 4.0 (see section 1 below)

The recommendations included in this report focus primarily on those areas where we saw gaps in the existing guidance – around licensing and metadata practices that support reuse, attribution and sharing. These recommendations are by nature respective of and only applicable to data that may be shared openly and without restrictions, per the allowances of any superseding national laws, security measures, and/or policies. For example, for National Meteorological and Hydrological Services who work with warnings, prohibit derivatives, etc., these recommendations may not apply.

We invite other climate data producers and publishers to adopt and provide input on these recommendations, especially fellow intergovernmental and government agencies and organizations that produce and publish climate datasets, data catalogs, and/or other data products.

PART 1: LEGAL AND LICENSING TERMS

Opening climate data requires an examination of the public's legal rights to access and use the climate data, often dictated by copyright and licensing (see definitions in Appendix). This legal detail is sometimes missing from data sharing and legal interoperability conversations. While existing policies and principles do discuss open sharing of data, and some mention the use of data licenses, our recommendations suggest two specific legal tools.

In order to make climate data easy to access and use legally, we have explained our two recommended options, in order for you to decide the best legal tool for sharing your climate data and climate database(s). We hope this adds helpful clarity and removes an often unintentional barrier to data use.

TWO RECOMMENDED OPTIONS

Option A		Option B
Creative Commons Public Domain Dedication (CC0 / "CC Zero") with Attribution Request	**or**	**Creative Commons Attribution 4.0 International License (CC BY 4.0)**

OPTION A: **CREATIVE COMMONS GLOBAL PUBLIC DOMAIN DEDICATION (CC0/"CC ZERO") WITH ATTRIBUTION REQUEST**

"This work is dedicated to the public domain, marked by CC0 1.0. Attribution is requested. To view a copy of this dedication, visit http://creativecommons.org/publicdomain/zero/1.0"

Who it's recommended for: Climate data organizations developing data policies for the first time; climate data producers and publishers that are willing to share their data freely in order to maximally reduce the friction in use.

What it is: CC0[4] is a dedication of the climate data to the worldwide public domain, and waives any existing or potential copyright and related rights, including database rights. CC0 is not a license; it is a public domain dedication. When CC0 is applied to a work, copyright no longer applies to the work in most jurisdictions around the world. In jurisdictions where a public domain dedication is not possible, CC0 provides a maximally permissive fallback

[4] CC0: https://creativecommons.org/share-your-work/public-domain/cc0/

license (see Appendix for specifics by jurisdiction). Our recommendation for "CC0 with Attribution Request" adds attribution practices by default, signaling the importance of giving credit through attribution when using climate data, even though attribution is not legally required for works in the public domain.[5]

What it allows users to do: "CC0 with Attribution Request" allows users worldwide to distribute, remix, adapt, and build upon the material in any medium or format, with no legal conditions on reuse. Because users always need to review the legal and licensing terms of their climate data sources and determine what's allowed, data already in the public domain will always be the easiest to use, redistribute, and/or remix. Alongside the use of CC0, attribution is requested in a standard format to allow reusers to easily provide credit to the climate data producer or publisher (standard attribution practice is to include, at a minimum, four key components with their respective links: Title (of the work), Author (author's name), Source (link to the original work - e.g., DOI or URL), and License or Public Domain (in this case "CC0 with Attribution Request").

Getting credit through attribution as a professional norm: "CC0 with Attribution Request" relies on climate data users to attribute (or give credit to) source climate data and databases. When using or referencing someone else's work, attribution is widely viewed as an expected and standard professional practice and driver of research and progress, rather than an action arising from legal enforcement and the fear of legal consequences. Attribution should be simple and always formatted to include four standard "TASL" elements: Title, Author or Publisher, Source Link, and Licence or Legal Terms (see Part 2 below for more information below about attribution formatting in metadata). Climate data providers can ensure they get credit and make it easy for users to provide attribution by providing sample attribution statements adjacent to the work, on your website, and/or as part of your community norms when they use your climate data.

- Sample text, used by UCLA Dataverse: "Our Community Norms as well as good scientific practices expect that proper credit is given via citation. Please use the data citation shown on the dataset page."

[5] Attribution to the original work is a legal requirement of CC licenses. It needs to include, at a minimum, four key components: the author's name, title of the work, the link to the original work (e.g., DOI, website url), and licensing information.

Pros - Maximized reuse and less legal friction: "CC0 with Attribution Request" enables sharing of climate data with the least amount of friction by removing legal uncertainty around reuse restrictions. Users do not need to concern themselves about the legality of their data reuse, saving time that would otherwise be spent deciphering, negotiating, and following complex legal terms and conditions. This legal tool also explicitly specifies disclaimers of warranty that many climate data providers need in order to distribute their data and reduces time spent paying lawyers to develop and negotiate custom end-user license agreements (EULAs).[6]

Considerations - Proactively managing for sensitive data and misuse: Organizations with particularly sensitive data can take steps to address any sensitivities before applying "CC0 with Attribution Request" once the data is available to share openly. Organizations may also consider tracking and regulating the use of their name trademarks.

OPTION B: **CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE (CC BY 4.0)**

"This work is licensed under CC BY 4.0. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/"

Who we recommend for: Climate data producers and publishers that want to share their data and retain ownership; producers and publishers of forms of climate data that are copyrightable in nearly all jurisdictions; organizations that have already shared their climate data under an open copyright license; climate data producers and publishers that choose to

[6] Many data providers want an explicit disclaimer of liability in order to avoid an implication that giving data away freely comes with any warranties or representations. In fact, one way to monetize freely-provided data can be where a reuser wants warranties and representations--the provider can negotiate for payment for taking on that liability. This would be uneconomical to offer for free, as the provider would be taking a financial risk that the reuser will make a claim about the correctness or utility of the data.

leverage data and database rights in jurisdictions that allow copyright on data and database rights[7] (see Appendix for specifics by jurisdiction).

What it is: In some jurisdictions, copyright holders can retain data and database rights through copyright and grant use of those copyrighted works to others using CC licenses. CC BY 4.0 is the most permissive Creative Commons license, allowing for unrestricted reuse as long as the work is appropriately attributed.

What it allows users to do: The CC BY 4.0 copyright license allows users to distribute, remix, adapt, and build upon the material in any medium or format, even commercially, so long as attribution is given to the data producer or publisher.

Getting credit through legal requirements: The difference between "CC0 with Attribution Request" and CC BY 4.0, in terms of receiving credit for the work, is that CC0 with Attribution Request requires attribution through professional norms, whereas CC BY 4.0 requires attribution through the terms of the open copyright license. With CC BY 4.0, organizations have legal means to pursue action in court for infringement of the license if attribution is not provided. There are cases under this method where attribution cannot be enforced: for example, where the data was not copyrightable in the first place, or where a reuser is making use of the data under exceptions and limitations to copyright such as fair use.

Pros - Property ownership and license enforcement: Where allowed, CC BY 4.0 permits organizations to retain their data as copyrighted property. CC BY 4.0 also creates a stronger mechanism for enforcing attribution through copyright infringement actions or other means such as website takedown requests. This legal enforcement of attribution may also allow you to more accurately track attributions as a way to determine the socio-economic benefits of your data provision. You can and should use supporting terms and conditions in order to meet any additional needs of your organization.

[7] Even where data may not be copyrightable, a data license (or "terms and conditions" or "end user license agreement") may create contractual limitations that govern the use of climate data. For climate data hosts in countries that allow data and databases rights, we still recommend the use of CC0 to dedicate the data and databases to the worldwide public domain. If they are not able to, we recommend the climate data and databases be openly licensed using the most current version of the Creative Commons Attribution International Copyright License (CC BY).

Considerations: Using CC BY 4.0 instead of "CC0 with Attribution Request" can make sharing harder down the line, e.g., when the openly licensed data is used to create a derived product by merging it with other openly licensed data; attribution to individual data authors and keeping track of changes to the original work (which must also be listed in the attribution statement) can become increasingly difficult to track as works are used and reused hundreds or thousands of times. Additionally, be sure to confirm your data and/or database is legally copyrightable in the first place as only some legal jurisdictions provide these rights. CC BY 4.0 is not applicable on material that is uncopyrightable material or being used under exceptions and limitations to copyright. In these cases, attribution may still be requested, just as in a CC0 with Attribution Request release.

THE REALITIES OF LICENSE STACKING AND ATTRIBUTION STACKING

Many uses of climate data require gathering data from multiple sources in order to combine them into a resulting derivative data product (or remix). Although we universally recommend all produced climate data use either CC0 with Attribution Request or CC BY 4.0, we acknowledge that currently there are many climate data sources that carry varying legal and licensing terms. We aim to specifically address this: working with multiple data sources can be tricky due to "license stacking," where remixing or reusing data has the potential to lead to 1) more restrictive licensing, 2) incompatible licenses, and 3) a complicated situation that requires a lawyer. It is important that users feel confident about picking a license for remix/reuse, and confident about navigating what can become complicated license stacking scenarios.

If you add a creative element to any pre-existing work (e.g. prediction modeling from a group of observation data), that addition makes you eligible to claim copyright over those new creative elements. Given the goal of these recommended best practices is to reduce friction for the field of climate data users, we recommend climate data producers and hosts seek to maximize the impact and reach of the work, and choose to license the data using the least

restrictive legal terms as possible, with "CC0 with Attribution Request" as default preference if applicable[8], and CC BY 4.0 as second preference if applicable.

In all, do what is needed to ensure you and your climate data sources have sufficient rights to release and share work under legal and licensing terms that are as open as possible. In general, you are required to maintain compliance with any legal restrictions included with your source data, and your resulting remixed products must be licensed accordingly.

To maximize open sharing of climate data products that have been created from sources with restricting legal terms, begin by trying to work with those sources to open their climate data using permissive legal tools. If the source data providers are unable or unwilling to dedicate their data and database rights to the public domain using CC0, the next best option is to work with the source data providers to openly license their data using the most current version of CC BY 4.0. Another option is to try to contractually procure the rights to use and redistribute it using either CC0 with Attribution Request or CC BY 4.0, covering the use of the source data, redistribution of the source data, and distribution of the derived product.[9]

Although these steps will take more work, they raise the rates of legal compliance and ensure you and future users can use and freely share the data. Alternatively, you might choose to apply a more restrictive license on them, and, depending on the license of the pre-existing work, you may be able to distribute the entire remix under the more restrictive terms. This is worth consideration when being mindful of any privacy information that might be included in your climate data.[10] We encourage collaboration to identify ways to protect privacy before sharing. This might include:

- Datasets that contain personally identifiable information (PII).

[8] Unless all data sources have been committed to the public domain, you will not be allowed to use CC0 with Attribution Request on a remix, as that would be violating and attempting to override the more-restrictive terms of your sources. If the existing terms and conditions or licenses prevent you from reusing and/or remixing the source data, work with the source data providers to adjust or consider procuring the rights to do so.
[9] Here is an example from NOAA: https://nauticalcharts.noaa.gov/data/data-licensing.html
[10] The U.S. Privacy Act addresses this: https://www.hhs.gov/foia/privacy/index.html#:~:text=The%20Privacy%20Act%20of%201974,other%20identifying%20number%20or%20symbol

- Datasets that contain information of relevance for national security interests.
- Review articles and other works of synthesis or opinion/analysis where grantees are invited to contribute on a specific topic.

There are many valid reasons to restrict data access, including: data that contains personal information, cases where consent has not been given for release, confidential commercial information, or situations where there are sound public reasons for restricting data (e.g. protection of endangered species or archaeological sites). Anonymization techniques, data sharing agreements, and safe havens where data can be accessed in controlled and secure circumstances (e.g., data trusts) can be useful in such cases. When specific legal or ethical restrictions prohibit public sharing of climate data, it is still important to indicate how others may obtain access to the data.

If your organization could benefit from support from Creative Commons in designing data sharing agreements, crafting open grant requirements, and/or providing training on CC legal tools, please let us know by emailing openclimatedata@creativecommons.org.

If you have directly published someone else's data for the purpose of hosting (without any alteration), you have not made any copyrightable changes and therefore cannot claim any additional copyright. In this case, you always inherit and apply the legal and licensing terms applied by the source. If publishing multiple external datasets to a catalog, aim to apply a universal set of legal terms to the full set of external data when possible. While this may not be possible to do if source data exists under incompatible licenses, it can be desirable to work with the source data provider and ensure that you are legally allowed to distribute all data under a standard set of terms. Additionally, you may request attribution as the publisher/host, but not as a matter of copyright.

In addition to license stacking, working with multiple data sources and their attribution requirements can also lead to "attribution stacking," where each remix and/or reuse of a dataset results in 1) a longer attribution statement and 2) more complicated attributions. This is because the attribution responsibility passes on in perpetuity. In other words, you are

required to attribute all of your sources and their sources (unless they are in the public domain or have decided not to request attribution), while users of your remix product are also required to attribute you, your sources, and their sources. We address this challenge with our recommendations in Part 2 below.

PART 2: METADATA VALUES FOR SHARING, ATTRIBUTION, AND PROVENANCE

Comprehensive, clear (plain language), and machine-readable metadata maximizes the reusability of climate data by enabling replication and integration across different contexts. It makes climate data findable and improves search-engine optimization (SEO) for federated search engines, as well as for your own organizational internal search engines. It boosts interoperability by providing qualified references to other (meta)data (see FAIR principles for more on this).

There are existing model metadata schemas and standards that exist[11,12], including those that address legal license and attribution information. With these recommendations, we address a gap and create and recommend **metadata values that specifically standardize the inclusion of upfront, clear values that provide information about attribution, legal and licensing terms, and provenance[13].** Metadata is the best place to be proactive about these critical aspects of better sharing.

---

[11] See the W3C DCAT specification: https://www.w3.org/TR/vocab-dcat/, ODI Open Data Rights Statement Vocabulary: https://schema.theodi.org/odrs/, and ISO complete list of projects: https://committee.iso.org/sites/tc211/home/projects/projects---complete-list.html

[12] GEO DMP-3: Data should be structured using encodings that are widely accepted in the target user community and aligned with organizational needs and observing methods, with preference given to non-proprietary international standards.

[13] We also recommend organizations reference ISO and DCAT metadata standards for additional guidance for provenance.

Both of our recommended options for legal and licensing terms – CC0 with Attribution Request and CC BY 4.0 – rely on attribution as a key component of legality and open sharing. At Creative Commons we've established a standard practice of using "TASL" formatting – Title, Author or Publisher, Source Link, and License – in all attribution statements. The title allows the user to visually identify the specific source being referenced, while the source link takes the user directly to that source (we strongly recommend the use of DOIs as first best option whenever possible), and license names "CC0 with Attribution Request" for the user's convenience. If possible, include hyperlinks as follows: Author or Publisher linking to the data catalog or website, Source linking to the DOI, and License linking to the Creative Commons reference page.

We recommend standard use of the following human-and machine-readable metadata values. These interoperable values support better data sharing by acknowledging the sources and future uses of climate data, catalogs, and derived products for better sharing. These values bring user-friendly transparency, reduce guesswork, simplify data discovery and comprehension, and ensure broader accessibility.

In cases of missing or unavailable metadata, it's important to include placeholders (e.g., "N/A") to promote the standard use of the additional recommended values. This also supports machine-readable guideline compliance and improved interoperability for using and sharing data in a way that can enable automation of information. In general, these recommended metadata values should be used in addition to any other values already used by your organization.

To start, these are our recommended metadata schemas and values, provided in the context of hypothetical examples of data published by the Korea Meteorological Administration that is being shared and used in different ways. Details and definitions of each recommended value are provided in the next section. In each of the example metadata schemas provided below, **we recommend the use of both of two distinct groups of attribution values: the first group provides guidance on how to attribute resulting data, while the second group offers information needed to attribute the sources of that data.** The second group can be repeated as needed for additional sources of data used, as demonstrated in our example schema below, where we used "Source 1" to indicate metadata pertaining to the first data

source in the series. This is especially useful for users to track the lineage or provenance of the climate data and understand if and how your data has been remixed from multiple sources.

Scenario A	Scenario B	Scenario C
Data is produced by the Korea Meteorological Administration (KMA)	Climate Data Store (CDS) hosts Korea Meteorological Administration (KMA) data in its catalog and attributes KMA	User remixes KMA data from CDS and data from the Japanese Meteorological Agency (JMA), and attributes CDS, KMA, JMA

Sample Scenario A

Metadata for **Korea Meteorological Administration (KMA) - Producer of Data**

Group	Metadata Value Label	Example Value
ABOUT MY DATA/ DATABASE/ PRODUCT *My organization has published something and made it available to share with you, and we want to make sure you*	**Title** (equivalent to **dcterms:title**) *The **title** of this work is ____.*	Korean Land and Water Levels 2003
	Publisher (equivalent to **dcterms:publisher**) *Please list us as the **publisher**.*	Korea Meteorological Administration (KMA)
	Identifier (equivalent to **dcterms:identifier**) *The **identifier**[14] for this data is*	doi:XXXXXXX/kma.domain

[14] While metadata standards are flexible with the type of value used for Identifier, we strongly recommend a default use of Digital Object Identifiers (DOIs) wherever possible. These should be used as standard machine-readable components of attribution statements in order to track provenance and understand where your data is being used. If you are the host of a data catalog that has published a copy of external data, you would use a new DOI that links to your catalog. However, links to external data may use their original DOIs.

attribute it to us and our sources.	*__(doi:XXXXXXX/XXXXXXXXXXXX)__.*	
	License (equivalent to **dcterms:license**, and formatted as a URL) *The **license**[15] for this data is __(URL)__.*	http://creativecommons.org/publicdomain/zero/1.0
	Rights Statement (equivalent to **dct:rightsStatement**) *Context: Here is a **statement that previews the licensing rights and legal terms**[16] for this data.*	CC0 with Attribution Request. This work is dedicated to the public domain, marked by CC0 1.0, which allows users worldwide to distribute, remix, adapt, and build upon the material in any medium or format, with no legal conditions on reuse. Attribution is requested in a standard format to allow reusers to include this information easily.
	Bibliographical Citation (equivalent to **dcterms:bibliographicCitation**)	Korean Land and Water Levels 2003, Korea Meteorological

[15] The label "license" is inherited from DCAT metadata standards and is used to also capture CC0 with Attribution Request as a URL linking to CC0 (http://creativecommons.org/publicdomain/zero/1.0), which is not technically a license but still sets legal terms.

[16] For CC0 with Attribution Request: "This work is dedicated to the public domain, marked by CC0, which allows users worldwide to distribute, remix, adapt, and build upon the material in any medium or format, with no legal conditions on reuse. Attribution is requested in a standard format to allow reusers to include this information easily." For CC BY 4.0: "This work is licensed under CC BY 4.0. The CC BY license allows users to distribute, remix, adapt, and build upon the material, even commercially, in any medium or format, so long as attribution is given to the data producer or publisher."

	*Here is how to **cite us and our sources**[17]:* • *[Our Citation: Title, Publisher, Identifier, License, etc.]* • *[Our Source 1: Title, Publisher, Identifier, License, etc.]* • *[Our Source 2: Title, Publisher, Identifier, License, etc.]* • *… etc.*	Administration (KMA), doi:XXXXXXX/kma.domain, http://creativecommons.org/publicdomain/zero/1.0
ABOUT MY SOURCE(S) *My organization got data from other sources, and we want to make sure we attribute our sources.*	**Source 1 Title** (**dcterms:source1Title**) *Context: The **title** of our first source[18] is ____.*	N/A
	Source 1 Publisher (**dcterms:source1Publisher**) *Context: our first source came from __(**publisher**)__.*	N/A
	Source 1 Identifier (**dcterms:source1Identifier**) *Context: The **identifier**[19] for our first source is __(doi:XXXXXXX/XXXXXXXXXXX)__.*	N/A
	Source 1 License (**dcterms:source1License**, and formatted as a URL)	N/A

[17] This metadata value provides the user a complete and copy-paste-ready attribution statement combined from the previous values provided above.

[18] Showing the name (or other identifying text) improves the user experience by allowing users, producers, and suppliers to visually confirm the data source they are linking to is correctly associated with the source data/catalog/product.

[19] Use Digital Object Identifiers (DOIs) as standard machine-readable components of attribution statements in order to track provenance and understand where your data is being used.

	*Context: The **license** for our first source is __(URL)__.*	
	Source 1 Bibliographical Citation (**dcterms:source1BibliographicCitation**) *Context: Here is the **citation of our Source 1 and their sources**:* • *[Our Source 1: Title, Publisher, Identifier, License, etc.]* • *[Source 1's Source 1: Title, Publisher, Identifier, License, etc.]* • *[Source 1's Source 2: Title, Publisher, Identifier, License, etc.]* • *…*	N/A

Sample Scenario B

Metadata for **Climate Data Store (CDS) - Host as a Catalog**

Group	Metadata Value Label	Example Value
ABOUT MY DATA/ DATABASE/ PRODUCT *My organization has published something and made it available to share with you, and we want to make sure you attribute it to us and our sources.*	**Title** (equivalent to **dcterms:title**) *The **title** of this work is ____.*	KMA Korean Land and Water Levels 2003
	Publisher (equivalent to **dcterms:publisher**) *Please list us as the **publisher**.*	Climate Data Store (CDS)
	Identifier (equivalent to **dcterms:identifier**)	doi:XXXXXXX/cds.domain

	*The **identifier**[20] for this data is __(doi:XXXXXXX/XXXXXXXXXXX)__.*	
	License (equivalent to **dcterms:license**, and formatted as a URL) *The **license**[21] for this data is __(URL)__.*	http://creativecommons.org/publicdomain/zero/1.0
	Rights Statement (equivalent to **dct:rightsStatement**) *Context: Here is a **statement that previews the licensing rights and legal terms**[22] for this data.*	CC0 with Attribution Request. This work is dedicated to the public domain, marked by CC0 1.0, which allows users worldwide to distribute, remix, adapt, and build upon the material in any medium or format, with no legal conditions on reuse. Attribution is requested in a standard format to allow reusers to include this information easily.

[20] While metadata standards are flexible with the type of value used for Identifier, we strongly recommend a default use of Digital Object Identifiers (DOIs) wherever possible. These should be used as standard machine-readable components of attribution statements in order to track provenance and understand where your data is being used. If you are the host of a data catalog that has published a copy of external data, you would use a new DOI that links to your catalog. However, links to external data may use their original DOIs.

[21] The label "license" is inherited from DCAT metadata standards and is used to also capture CC0 with Attribution Request as a URL linking to CC0 (http://creativecommons.org/publicdomain/zero/1.0), which is not technically a license but still sets legal terms.

[22] For CC0 with Attribution Request: "This work is dedicated to the public domain, marked by CC0, which allows users worldwide to distribute, remix, adapt, and build upon the material in any medium or format, with no legal conditions on reuse. Attribution is requested in a standard format to allow reusers to include this information easily." For CC BY 4.0: "This work is licensed under CC BY 4.0. The CC BY license allows users to distribute, remix, adapt, and build upon the material, even commercially, in any medium or format, so long as attribution is given to the data producer or publisher."

	Bibliographical Citation (equivalent to **dcterms:bibliographicCitation**) *Here is how to **cite us and our sources**[23]:* • *[Our Citation: Title, Publisher, Identifier, License, etc.]* • *[Our Source 1: Title, Publisher, Identifier, License, etc.]* • *[Our Source 2: Title, Publisher, Identifier, License, etc.]* • *… etc.*	KMA Korean Land and Water Levels 2003, Climate Data Store, doi:XXXXXXX/cds.domain, http://creativecommons.org/publicdomain/zero/1.0, CC0 with Attribution Request ; Korean Land and Water Levels 2003, Korea Meteorological Administration, doi:XXXXXXX/kma.domain, http://creativecommons.org/publicdomain/zero/1.0, CC0 with Attribution Request
ABOUT MY SOURCE(S) *My organization got data from other sources, and we want to make sure we attribute our sources.*	**Source 1 Title** (**dcterms:source1Title**) *Context: The **title** of our first source[24] is ____.*	Korean Land and Water Levels 2003
	Source 1 Publisher (**dcterms:source1Publisher**) *Context: our first source came from __(**publisher**)__.*	Korea Meteorological Administration
	Source 1 Identifier (**dcterms:source1Identifier**)	doi:XXXXXXX/kma.domain

[23] This metadata value provides the user a complete and copy-paste-ready attribution statement combined from the previous values provided above.

[24] Showing the name (or other identifying text) improves the user experience by allowing users, producers, and suppliers to visually confirm the data source they are linking to is correctly associated with the source data/catalog/product.

	*Context: The **identifier**[25] for our first source is __(doi:XXXXXXX/XXXXXXXXXXX)__.*	
	Source 1 License (**dcterms:source1License**, and formatted as a URL) *Context: The **license** for our first source is __(URL)__.*	CC0 with Attribution Request
	Source 1 Bibliographical Citation (**dcterms:source1BibliographicCitation**) *Context: Here is the **citation of our Source 1 and their sources**:* • *[Our Source 1: Title, Publisher, Identifier, License, etc.]* • *[Source 1's Source 1: Title, Publisher, Identifier, License, etc.]* • *[Source 1's Source 2: Title, Publisher, Identifier, License, etc.]* • *…*	Korea Meteorological Administration, Korean Land and Water Levels 2003, 01/01/2004, doi:XXXXXXX/kma.domain, CC0 with Attribution Request

Sample Scenario C Metadata for **User - Remixing Two Datasets into a New Product**		
Group	**Metadata Value**	**Example Value**
ABOUT MY DATA/ DATABASE/	**Title** (equivalent to **dcterms:title**)	User's Org

[25] Use Digital Object Identifiers (DOIs) as standard machine-readable components of attribution statements in order to track provenance and understand where your data is being used.

PRODUCT ***My organization has something to share with you, and we want to make sure you attribute us and our sources.***	*The **title** of this work is ____.*	
	Publisher (equivalent to **dcterms:publisher**) *Please list us as the **publisher**.*	Land and Water Levels Across Asia 2003
	Identifier (equivalent to **dcterms:identifier**) *The **identifier**[26] for this data is __(doi:XXXXXXX/XXXXXXXXXXX)__.*	doi:XXXXXXX/user.domain
	License (equivalent to **dcterms:license**, and formatted as a URL) *The **license**[27] for this data is __(URL)__.*	CC BY 4.0
	Rights Statement (equivalent to **dct:rightsStatement**) *Context: Here is a **statement that previews the licensing rights and legal terms**[28] for this data.*	This work is licensed under CC BY 4.0, which allows users to distribute, remix, adapt, and build upon the material, even commercially, in any medium or format, so long as

[26] While metadata standards are flexible with the type of value used for Identifier, we strongly recommend a default use of Digital Object Identifiers (DOIs) wherever possible. These should be used as standard machine-readable components of attribution statements in order to track provenance and understand where your data is being used. If you are the host of a data catalog that has published a copy of external data, you would use a new DOI that links to your catalog. However, links to external data may use their original DOIs.

[27] The label "license" is inherited from DCAT metadata standards and is used to also capture CC0 with Attribution Request as a URL linking to CC0 (http://creativecommons.org/publicdomain/zero/1.0), which is not technically a license but still sets legal terms.

[28] For CC0 with Attribution Request: "This work is dedicated to the public domain, marked by CC0, which allows users worldwide to distribute, remix, adapt, and build upon the material in any medium or format, with no legal conditions on reuse. Attribution is requested in a standard format to allow reusers to include this information easily." For CC BY 4.0: "This work is licensed under CC BY 4.0. The CC BY license allows users to distribute, remix, adapt, and build upon the material, even commercially, in any medium or format, so long as attribution is given to the data producer or publisher."

		attribution is given to the data producer or publisher.
	Bibliographical Citation (equivalent to **dcterms:bibliographicCitation**) *Here is how to **cite us and our sources**[29]:* • *[Our Citation: Title, Publisher, Identifier, License, etc.]* • *[Our Source 1: Title, Publisher, Identifier, License, etc.]* • *[Our Source 2: Title, Publisher, Identifier, License, etc.]* • *… etc.*	User's Org, Land and Water Levels Across Asia 2003, 31/10/2024, doi:XXXXXXX/user.domain, CC BY 4.0. ; Climate Data Store (CDS), KMA Korean Land and Water Levels 2003, 21/05/2004, doi:XXXXXXX/cds.domain, CC0 with Attribution Request ; Korea Meteorological Administration, Korean Land and Water Levels 2003, 01/01/2004, doi:XXXXXXX/kma.domain, CC0 with Attribution Request ; Japan Meteorological Agency, Japanese Land and Water Levels 2003, 01/01/2004, doi:XXXXXXX/jma.domain, CC BY 4.0.
ABOUT MY SOURCE(S)	**Source 1 Title** (**dcterms:source1Title**) *Context: The **title** of our first source[30] is ____.*	KMA Korean Land and Water Levels 2003

[29] This metadata value provides the user a complete and copy-paste-ready attribution statement combined from the previous values provided above.

[30] Showing the name (or other identifying text) improves the user experience by allowing users, producers, and suppliers to visually confirm the data source they are linking to is correctly associated with the source data/catalog/product.

My organization got data from other sources, and we want to attribute our sources.	**Source 1 Publisher** (**dcterms:source1Publisher**) *Context: our first source came from __(**publisher**)__.*	Climate Data Store (CDS)
	Source 1 Identifier (**dcterms:source1Identifier**) *Context: The **identifier**[31] for our first source is __(doi:XXXXXXX/XXXXXXXXXXX)__.*	doi:XXXXXXX/cds.domain
	Source 1 License (**dcterms:source1License**, and formatted as a URL) *Context: The **license** for our first source is __(URL)__.*	CC0 with Attribution Request
	Source 1 Bibliographical Citation (**dcterms:source1BibliographicCitation**) *Context: Here is the **citation of our Source 1 and their sources**:* • *[Our Source 1: Title, Publisher, Identifier, License, etc.]* • *[Source 1's Source 1: Title, Publisher, Identifier, License, etc.]* • *[Source 1's Source 2: Title, Publisher, Identifier, License, etc.]* • *…*	Climate Data Store (CDS), "KMA Korean Land and Water Levels 2003", 21/05/2004, doi:XXXXXXX/cds.domain, CC0 with Attribution Request ; Korea Meteorological Administration, "Korean Land and Water Levels 2003", 01/01/2004, doi:XXXXXXX/kma.domain, CC0 with Attribution Request

[31] Use Digital Object Identifiers (DOIs) as standard machine-readable components of attribution statements in order to track provenance and understand where your data is being used.

	Source 2 Title (**dcterms:source2Title**) *Context: The **title** of our second source[32] is ____.*	Japanese Land and Water Levels 2003
	Source 2 Publisher (**dcterms:source2Publisher**) *Context: our second source came from __(**publisher**)__.*	Japan Meteorological Agency (JMA)
	Source 2 Identifier (**dcterms:source2Identifier**) *Context: The **identifier**[33] for our second source is __(doi:XXXXXXX/XXXXXXXXXXX)__.*	doi:XXXXXXX/jma.domain
	Source 2 License (**dcterms:source2License**, and formatted as a URL) *Context: The **license** for our second source is __(URL)__.*	CC BY 4.0
	Source 2 Bibliographical Citation (**dcterms:source2BibliographicCitation**) *Context: Here is the **citation of our Source 2 and their sources**:*	Japan Meteorological Agency, Japanese Land and Water Levels 2003, 01/01/2004, doi:XXXXXXX/jma.domain, CC BY 4.0

[32] Showing the name (or other identifying text) improves the user experience by allowing users, producers, and suppliers to visually confirm the data source they are linking to is correctly associated with the source data/catalog/product.

[33] Use Digital Object Identifiers (DOIs) as standard machine-readable components of attribution statements in order to track provenance and understand where your data is being used.

	• *[Our Source 2: Title, Publisher, Identifier, License, etc.]* • *[Source 2's Source 1: Title, Publisher, Identifier, License, etc.]* • *[Source 2's Source 2: Title, Publisher, Identifier, License, etc.]* • *…*	

PART 3: GENERAL MANAGEMENT AND GOVERNANCE ACTIONS

ASSIGN A TECHNICAL MANAGING STEWARD

- Designate a responsible party for managing datasets, sourcing, attributions, Digital Object Identifiers (DOIs), and their other associated metadata. This responsibility goes beyond ensuring a user-friendly interface; it entails the maintenance of accurate and aligned metadata with the data itself. The accuracy of metadata is essential for maintaining reliability.

ASSIGN A LEGAL AND/OR POLICY STEWARD

- Designate a responsible party for ensuring governing policy guidelines are adopted and implemented. This includes the recommendations offered herein, alongside any other governance policies that apply. This work is often done in alignment with management, legal, and governing committees and boards, in ways that support the work of the technical managing stewards mentioned above.

SHARE IN PARTNERSHIP

- Proactively identify 2+ organizations/entities to share data with in partnership.
- Ask each partner what their barriers are, and co-create new sharing incentives that meet their needs.
- Intend large datasets for reanalysis, in partnership with related organizations/entities.

REVISIT AND UPDATE REGULARLY

- Revisit and update your data sharing policies frequently to coincide with evolving practices, technology, and needs.
- Regularly consult these recommendations and the FAIR assessment to assess and align internal data as it's added.

APPENDIX

DEFINITIONS

Copyright Terms

Copyright: Copyright grants to creators a bundle of exclusive rights over their creative works, which generally include, at a minimum, the right to reproduce, distribute, display, and make adaptations. The exact boundaries of copyright vary country by country, but as there is international harmonization, all implementations should contain these rights. Note that copyright does not apply to non-creative works, which may include some data sets or other material such as mathematical equations.

Licensing: Licensing gives copyright holders the space to specify the terms for permitted use of their copyrighted material. Licenses are operative only when applied to material in which a copyright exists, and even then only when a particular use would otherwise not be permitted by copyright.

Attribution (vs. Citation): Attribution to the original work is a legal requirement of CC licenses. It needs to include, at a minimum, four key components with their respective links: Title (of the work), Author (author's name), Source (link to the original work - e.g., DOI or URL), and License or Public Domain. These attributions can be and are often provided via citations, which typically already include the former two components: the title of the work

and the author's name; adding source links and license information to citations ensures that you have well-written citations which also meet the legal requirements of attribution.

Role Types

Data Producer/Supplier: The data producer or supplier is responsible for creating and sharing the data. They generate the data, conduct research to collect and compile climate data, and are the originators of the information. Data producers, like Sensor.Community, have direct influence on how the data they compile is made available to the public.

Data Host/Publisher: The data host or publisher holds the data for others and is responsible for making the data available to the public. They may or may not be the creators of the data, but they play a crucial role in facilitating data accessibility. Publishers, such as the Global Earth Observation System of Systems "GEOSS" Portal, have taken on the responsibility of making climate data available to the public by publishing it online, not necessarily exclusive of data sourced by other organizations or initiatives. Publishers have direct influence on how the data they publish is made available to the public.

Sponsors: Sponsors are organizations or entities that invest funds and resources in research for the collection and publication of climate data. They may have a direct influence on the accessibility of data resulting from sponsored research, even if the data is sourced by another organization.

> It is possible for an organization to play more than one of these three roles. For example, the National Oceanic and Atmospheric Administration (NOAA) is a U.S. Federal Government agency works as a sponsor, publisher, and producer of climate data, within its own agency and externally (e.g. the University of Michigan's Great Lakes Environmental Research Laboratory).

Source: The data source is where the data is most directly obtained and serves as a reference for metadata purposes. It indicates one step upstream in the data lineage history, which may or may not be the data producer. The data producer, on the other hand, refers to the origin or provenance of the data lineage history.

Data Types

Climate Data: Climate data on which we focused our recommendations align with the WMO Unified Data Policy domain of earth system data, which includes: weather, climate, hydrology, atmospheric composition, cryosphere, oceans, and space weather. These data represent the Earth system's past, present, and future states, exchanged in real or near-real time, and from historical sources. Data types include observations, satellite data, GIS data, video/photo data, and derived products (e.g. weather forecast and hydrological model.)

COPYRIGHT BY JURISDICTION

In the U.S., most data are considered discoverable "facts," and not original works, and are thus not copyrightable, meaning they cannot be owned by any entity. There are some types of climate data, such as satellite imagery, which may qualify for copyright protection in the U.S. although they also are primarily factual knowledge. Regardless, data or other works produced by the U.S. Government are generally not subject to copyright protection in the United States by law (17 USC 105(a)).

- In the U.S., databases are only copyrightable under certain conditions, specifically when the arrangement and selection of data are sufficiently creative and/or original.

- In the E.U., data are considered content of databases and can be protected under the sui generis database right under certain conditions if the maker is an EU resident and has made a substantial investment in obtaining, verifying, or presenting the data contained in a database. The sui generis database right is unrelated to copyright.

- In the E.U., databases as structures are copyrightable if they are original intellectual creations. (This is similar to the U.S. treatment of database copyright.) Databases with eligible content may also have their content protected by sui generis database rights (see above); these are two distinct sets of rights.

- In Canada, databases are only copyrightable under certain conditions, specifically when the arrangement and selection of data are sufficiently creative and/or original. The standard for copyrightability of databases can be different than in the U.S., and may grant copyright to some database works that the U.S. would not see as copyrightable.

REFERENCES

- FAQs about CC0
- FAQ about CC licenses
- The European Commission adopted CC0 and CC BY, after a fairly detailed report by the EC JRC Central IP Service
- Here is an example: How the Harvard Dataverse uses CC0 with citations
- A description from Dan Cohen at Northeastern University
- How open data publishing platform & community Dryad uses CC0
- How the dblp computer science facility at Dagstuhl Institute of Trier University switched to the use of CC0.
- Longstanding U.S. government guidance recommends CC0
- Norway uses CC BY
- ECMWF uses CC BY on a subset of real-time forecast data
- EUMETSAT's open data is free and unrestricted in accordance with WMO data policy.
    - WMO Unified Data Policy Annex 1
- Revised GEO Data Sharing and Data Management Principles. GEO does not make specific recommendations on legal interoperability.
    - Resource: GEO Statement on Open Knowledge
- Current Uses of CC0
    - Orgs using CC0 for data
    - KNMI (Netherlands) released their observation data under CC0, although their weather forecast data and climate data use different licenses