Neotoma Data Policy: v.2025

Status

Approved

Versions

V.2025. Current version
This version was reviewed and approved by the Neotoma Leadership Council on March 27, 2025.

V.2019. Prior version, now deprecated.

Policy Text

1. Overview and General Principles

1.1 Neotoma Data Governance

The Neotoma Paleoecology Database is an international coalition of multiple Constituent Databases. Neotoma strives to follow best practices in ethical open science, so that the repository continually works towards better alignment with TRUST[1], FAIR[2], and CARE[3] principles and satisfying public access mandates across the variety of international agencies that fund much of the science underlying Neotoma datasets.

Because many different kinds of paleoecological and paleoenvironmental data are stored in Neotoma, each corresponding to different communities of experts, Neotoma follows a system of centralized IT and distributed scientific governance. The Constituent Databases of Neotoma and their Stewards play a primary role in making decisions about how data should be added and curated.

Neotoma's Stewards and Standards Committee and Executive Committee help adjudicate data and governance issues that are beyond the scope of a single Constituent Database, while the Neotoma Leadership Council has the final say on policy matters and data-related decisions. See the Neotoma Bylaws for further information.

---

[1] https://doi.org/10.1038/s41597-020-0486-7
[2] https://doi.org/10.1038/sdata.2016.18
[3] https://doi.org/10.5334/dsj-2020-042

1.2 Scope of the Data Policy

The Data Policy described here is the default for Neotoma and all of its Constituent Databases. Individual Constituent Databases may create variations or more specific implementations of this Data Policy, as long as these modifications do not violate Neotoma's Creative Commons license, are clearly documented (see Appendix 2), and implementation is technically feasible.

1.3 Neotoma Principles of Data Access

1. Neotoma is an open-access data repository and research resource. The default expectation is that data in Neotoma are available to anyone.
2. Neotoma always strives to enhance data accessibility and make data available to as many audiences as possible, within the constraints of resource availability, by A) making data available in multiple forms and through multiple open software services and B) supporting the training and empowerment of data users, contributors, and stewards.
3. Some datasets in Neotoma may contain sensitive data and require additional protections. Neotoma currently implements some data protections through geolocational fuzzing, and is developing infrastructure to implement other data protections through embargoes. Upon implementation, it will be possible to request and obtain data embargoes under certain specific conditions (see Section 4.3 for more information).
4. These data protections follow from two primary intents:
    a. One intent is to enable the early submission of data to Neotoma prior to publication, while protecting the right of data generators and data synthesizers to first use of the uploaded data. This enables scientists to better follow best practices, in which data and metadata are jointly managed in start-to-finish scientific workflows. In this case, data embargoes are inherently temporary, with data publicly accessible upon publication.
    b. A second intent is to limit access to all or part of a data record that may require differential access due to data sensitivity, e.g. for sites vulnerable to destruction or Indigenous governance. In this case, a long-term embargo or geolocational fuzzing may be established.

2. Data Use

All data available through the Neotoma Paleoecology Database are free to use with attribution under a CC BY 4.0 license [link].

3. Data Citation and Attribution

Complete attribution operates at three levels: 1) a reference to the Neotoma Paleoecology Database (Appendix 1); 2) reference to the relevant Constituent Database(s) (Appendix 2); and 3) references to all original investigators and primary publications for the data used.

Any publication using Neotoma data should include a data table in the main text, or supplementary information, or external permanent archive that lists each site or dataset used, cites the associated primary publication, names the Constituent Database that curates this dataset, and provides the associated Neotoma dataset DOI. Neotoma recommends that scientific journals and citation services include these citations and DOIs in their citation indices, so that the work of primary data contributors is properly acknowledged.

Appendix 1 holds the preferred citation and permanent identifier (PID) for Neotoma, standard text for the Acknowledgments, and information about how to build a data table that lists site metadata, publications, and constituent databases. Appendix 2 describes the preferred citation for individual Constituent Databases and any modifications to the Data Policy for the Constituent Database.

Upon publication, please send a copy of the published work or a link to the electronic resource to Neotoma, or to the leaders and data stewards for the relevant Constituent Database. See Section 7 for contact information. Your assistance helps document database usage.

4. Protections for Sensitive Data

Note: Creating a process for protecting scientific data involves both policy updates and technical implementation. This policy is being developed in advance of and as guidance to technical implementation of an embargo management system for Neotoma data, which is currently underway. At the time of writing (Nov, 2024), data protections supported by Neotoma include geolocational fuzzing. Policies describing embargo management will be enacted upon technical implementation of embargoes, and likely will be modified somewhat as the technical implementation finalizes.

4.1 Overview: Open Data Ethics and Sensitive Data

Although the norm for Neotoma is that all data are open, in some situations the ethical value of data openness must be balanced against other ethics. These include protecting the first-publication rights of research teams gathering new data or building new curated data syntheses, protecting sites at risk of exploitation, protecting the authority of Indigenous peoples to govern access to their data, or honoring legacy data agreements. Each is described more fully below, and other situations may exist; this list is not intended to be comprehensive.

1) ***Protecting first-publication rights for research teams***. Neotoma follows the standard ethic in academia that scientific teams conducting original research have the first right to publication of their work. Hence, data contributors may embargo access to any associated datasets stored in Neotoma prior to publication. This protection extends both to primary datasets collected by the research team and to data syntheses, in which the research team has found, obtained, curated, and uploaded data to Neotoma.

    Example 1: A team of zooarchaeologists and archaeobotanists begin a new dig, collect a series of radiocarbon and OSL dates, conduct palynological analyses, and identify a series of vertebrate fossils at the site. A pre-publication embargo is requested and provided for the geochronological, pollen, and vertebrate datasets.

    Example 2: A research team is producing a synthesis of megafaunal extinction timing on Pacific Islands. They assemble a series of published geochronological dates on vertebrate specimens and upload these datasets to Neotoma. These datasets were not previously in Neotoma. A pre-publication embargo is requested and provided for the geochronological and vertebrate datasets at each site.

2) ***Protecting sites at risk of exploitation or destruction*.** Some sites and some kinds of data (e.g. fossils of charismatic megafauna, archaeological sites) are at risk of unauthorized access, exploitation, and destruction if their location is widely known. For these sites, data contributors or stewards may fuzz the location of the site and collection units within that site.
    a) Example 3: A bone bed is discovered on USA federal lands with a rich record of extinct ground sloths and other megafauna. To discourage unauthorized collections and record loss at this location, the permitting agency requests that site coordinates not be revealed. The site coordinates are fuzzed by a Neotoma Data Steward at time of upload.
3) ***Protecting the authority of Indigenous peoples to govern access to their data***. Neotoma recognizes the right of Indigenous communities to govern access to their data. The scope and implementation of Indigenous data governance is currently an active area of discussion, and the nature of these rights may vary by Indigenous community, country, and treaty. Neotoma takes a broad interpretation of these rights while still honoring its primary ethic of data openness.
    a) Example 3: A USA-based research team has engaged with tribal authorities about their lake sediment research on sites located on tribal lands as required by the terms of their National Science Foundation funding[4]. The tribal nation has approved the originally proposed research, but would like to approve new lines of research using these data by the original PIs or other researchers. In this case, a permanent embargo is requested for the datasets at the site and approved by the lead steward for the handling Constituent Database. Publicly-available metadata indicates the presence of pollen and geochronological datasets at the site, but

[4] See Chapter II.E.10, NSF 24-1 PAPPG: https://new.nsf.gov/policies/pappg/24-1

researchers will need to contact tribal authorities for permission to use the data. Once permission is granted and documented, the lead steward for the Constituent Database will share with the researcher the embargo keys for the full datasets.

4) ***Honoring legacy data contribution agreements.*** Some Constituent Databases were formed prior to the launch of Neotoma in 2010 and include legacy agreements for datasets that were provided to the database with access restrictions. This use case is supported by Neotoma, but it applies to a very small portion of Neotoma data holdings and is not a permanent solution. Constituent Databases with legacy data embargoes are responsible for finding a solution to sunset these embargoes.
    a) Example 4: A researcher contributed data to a large data synthesis effort in 1995 under the provision that their dataset was available upon request to the original data generator. No end-of-agreement timeline was included. Publicly-available metadata indicates the presence of pollen and geochronological datasets at a location, but researchers need to contact the embargo holder to obtain the authorization to view the full dataset.

4.2. Geolocational Fuzzing

For situations in which revealing the precise site coordinates might lead to destruction or exploitation of a site and its fossil and/or archaeological record, Neotoma will not store the precise coordinates of a site and collection units within that site. Instead, Neotoma will store geolocational information for a bounding rectangle that encompasses the site but is not centered on the site. Whether to employ a bounding rectangle instead of precise coordinates and the size of the rectangle is left to the decision of Constituent Databases and their Stewards.

4.3. Data Embargos

4.3.1 Overview

Access to and use of embargoed data requires permission from the original investigator or data synthesizer. Most embargoes of Neotoma datasets will be temporary. Embargoing is not automatic; embargoes must be requested by the data contributor(s) or placed by a Data Steward.

As noted above (Section 4), embargoes are not yet implemented within Neotoma; this development work is currently underway.

4.3.2 Technical Implementation

Embargoes are enforced at the level of dataset. For embargoed datasets, basic metadata are publicly viewable, but the datasets themselves are not. Basic metadata include geospatial coordinates, site name, dataset type, current end date of embargo, contact information for the

embargo holder, and other contact names as necessary (such as researcher names, if different from the embargo holder). Access will be enabled via an embargo key or similar solution.

Standard tools for searching the Neotoma Database (e.g. Neotoma Explorer, APIs, neotoma2 R package) can return basic metadata for embargoed sites, to the extent that it is visible per the embargo. For example, if a geochronological dataset for a site is embargoed, then temporal coordinates or other temporal information will not be returned for that site. Similarly, taxonomic information and data values associated with each taxon will not be returned for embargoed fossil datasets. Embargoed data will not be included in publicly posted full snapshots of the Neotoma Database. However, if an embargo is placed on data that were previously open, these data may be available in past snapshots.

4.3.3 Embargo Timing and Duration

Most embargoes are temporary and last for a defined time period, specified by the user at the time of embargo up to a pre-set maximum. Normally, an embargo lasts two years after a dataset is uploaded to Neotoma, or until publication, whichever happens first.

1. Neotoma follows the Joint Data Archiving Policy (JDAP), in which data are normally made publicly available at time of publication.
2. The embargo access manager (see next section) is responsible for removing the embargo at time of publication. For datasets that have been published and for which no other sensitive-data protections exist, a Steward of the governing Constituent Database may remove an embargo at any time after publication.
3. Most embargoes will be automatically lifted after two years, unless the data contributor(s) request an extension. Up to two two-year extensions, or a longer-term embargo in the case of protections for sensitive data (Section 4.1), can be requested.

4.3.4 Embargoed Data Access: Access Managers and Stewards

Every embargoed dataset will have one or more access managers. Access managers and their designees can access their embargoed data in Neotoma, using standard tools such as Neotoma Explorer, Tilia, and APIs. Access managers also have authority to share the embargoed data with any other person or entity of their choosing.

Designation of access manager(s) will depend on use case. In most cases, the access manager will be the original data generator(s) or data uploader(s), e.g. the principal investigator, student, and/or analyst responsible for the data. For some kinds of sensitive data, the access manager may not be the original data generator or data synthesizer, but is another relevant entity with authority to control access to the dataset.

Stewards for the Constituent Database that curates the embargoed data may access and curate the data as part of their normal activities as Data Stewards, but may not share the embargoed

data with others nor use the embargoed data for their own research purposes, unless they have written permission from the access manager. This rule also applies to Neotoma database administrators.

4.3.5 Requesting an Embargo

[This section will be developed once embargo manager is developed]

5. Expectations and Ethics for Neotoma Data Users, Contributors, and Stewards

5.1 Co-Authorship

Normal ethics apply to co-authorship of scientific publications. Paleoecological datasets are labor-intensive and complex: they take years to generate and always have additional attributes and metadata not captured in Neotoma. Neotoma data quality also relies on expert curation by data stewards, each associated with one or more Constituent Databases. Although not required under terms of Neotoma's CC BY data license, users of data stored in Neotoma's Constituent Databases should consider inviting the original data contributor, or Constituent Database steward(s), to be a co-author(s) of any resultant publications if that contributor's data are a major portion of the dataset analyzed, or if a data contributor or steward makes a significant contribution to the analysis of the data or to the interpretation of results. For large-scale studies using many Neotoma records where co-authorship is not practical, possible, or reasonable, users should ensure that they include full information for all sites and datasets included in the study, including citations, following practices described in Section 3. Under no circumstance should authorship be attributed to data contributors or stewards, individually or collectively, without their explicit consent. Appropriate data citation and attribution practices are outlined in Section 3 and the Appendices.

5.2 Data Contributors

Neotoma generally trusts that the datasets contributed to it are in compliance with national laws and funder policies and follow disciplinary ethical standards. Because datasets in Neotoma are collected and then contributed to the database from researchers around the world, who are subject to a variety of different data access policies, it is always the responsibility of the PI of the team contributing the data to ensure that the contributed data complies with legal, regulatory, and disciplinary ethical standards. Neotoma recommends that Data Stewards check with the PI at time of data submission to confirm that this responsibility has been met. A data intake checklist is being developed to aid in this effort.

For example, as of July 2023, the USA NSF EAR Data and Sample Policy requires that all data collected or generated with EAR funding is publicly accessible immediately upon publication or within two years after completion of data collection or generation. Thus, if data are contributed to Neotoma and its Constituent Databases under an embargo, it is the PIs responsibility to manage emplacement and/or request removal of the dataset embargo to comply with these timelines, regardless of whether Neotoma's Data Policy permits embargo periods that persist beyond these timelines.

5.3 Stewards

Stewards are expected to help uphold the ethical standards and data policies of Neotoma. Because rules and norms vary among countries and disciplines, Stewards are best positioned to mediate between the Neotoma policies, expectations, and ethics as described here and those of their local context. Areas of substantive disagreement or concern between Neotoma data policy and local context should be brought to the attention of the Neotoma Stewards and Standards Committee or the Neotoma Executive Committee, with ultimate adjudication by the Neotoma Leadership Council.

6. Data Disagreements

If a disagreement arises over the proper handling of, data protections for, and attribution of data submitted to or stored within Neotoma, a resolution to this disagreement should first be sought within the Constituent Database charged with curating that data. The Lead Steward(s) of the Constituent Database should be the primary mediator of this resolution.

If a resolution cannot be reached at the level of a Constituent Database, or the resolution is appealed, the issue can be escalated by sending a written request to the Stewards and Standards Committee. The Stewards and Standards Committee will then inform the Executive Committee that an issue has been raised, review the matter, and provide a recommendation to the Executive Committee. The Executive Committee will then adjudicate the issue or, for particularly substantive matters, bring the issue to the Neotoma Leadership Council for discussion and adjudication. The decision of the Neotoma Leadership Council is final.

7. Contact Information

Constituent Database Leads: https://www.neotomadb.org/data/constituent-databases

Neotoma Executive Committee: EC membership and contact information is listed here. Alternatively, to reach all members, you can contact neotoma-contact@googlegroups.com

Neotoma Leadership Council: Leadership Council membership and contact information is listed here.

Stewards and Standards Committee: Neotoma-data-standards@googlegroups.com
[need to list somewhere]

Appendix 1. How to Cite the Neotoma Paleoecology Database

Citation

The preferred citation for the Neotoma Paleoecology Database is Williams, Grimm et al. (2018):

Williams, J.W., Grimm, E.G., Blois, J., Charles, D.F., Davis, E., Goring, S.J., Graham, R., Smith, A.J., Anderson, M., Arroyo-Cabrales, J., Ashworth, A.C., Betancourt, J.L., Bills, B.W., Booth, R.K., Buckland, P., Curry, B., Giesecke, T., Hausmann, S., Jackson, S.T., Latorre, C., Nichols, J., Purdum, T., Roth, R.E., Stryker, M., Takahara, H., 2018. The Neotoma Paleoecology Database: A multi-proxy, international community-curated data resource. Quaternary Research 89, 156-177.

Neotoma's Permanent identifier (PID): http://doi.org/10.17616/R3PD38

Acknowledgments

Any use of Neotoma data should include the following or similar text in the acknowledgments: *"Data were obtained from the Neotoma Paleoecology Database (http://www.neotomadb.org, http://doi.org/10.17616/R3PD38) and its constituent database(s) [list here[5]]. The work of data contributors, data stewards, and the Neotoma community is gratefully acknowledged."*

Supplementary Data Table

Neotoma site metadata, publications, and dataset DOIs can be retrieved using the Neotoma APIs, neotoma2 R package, or third-party applications such as FossilPol.

Appendix 2. Preferred Citation, Acknowledgments, and Data Use Policies for Constituent Databases

Constituent Databases are listed alphabetically. If a Constituent Database is not listed below, please use the standard Neotoma citation and policy above.

[5] Note: the Constituent Databases that curate the data used should be mentioned by name in the Acknowledgments, in addition to any citation requirements listed in Appendix 2.

Alpine Pollen Database (ALPADABA)

Please see the European Pollen Database section for Citation information.

BugsCEP

Citation

Please reference the use of BugsCEP with the following two citations, and include the date of download and program versions where data or methods are used:

> *Buckland, P.I. & Buckland, P.C. (2006). BugsCEP Coleopteran Ecology Package. IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series # 2006-116. NOAA/NCDC Paleoclimatology Program, Boulder CO, USA. URL:http://www.ncdc.noaa.gov/paleo/insect.html*
>
> *Buckland, P.I. (2007). "The Development and Implementation of Software for Palaeoenvironmental and Palaeoclimatological Research: The Bugs Coleopteran Ecology Package (BugsCEP)". PhD thesis, Environmental Archaeology Lab., Department of Archaeology & Sámi Studies. University of Umeå, Sweden. Archaeology and Environment 23, 236 pp + CD. Available online: http://www.diva-portal.org/umu/abstract.xsql?dbid=1105*

European Pollen Database (EPD)

Citation

The currently preferred citation for EPD (including ALPADABA) is Fyfe et al. (2009). Please also cite Giesecke et al. (2014) if extensive use is made of the chronologies described there.

> *Fyfe, R. M., J. L. de Beaulieu, H. Binney, R. H. W. Bradshaw, S. Brewer, A. Le Flao, W. Finsinger, M. J. Gaillard, T. Giesecke, G. Gil-Romera, E. C. Grimm, B. Huntley, P. Kunes, N. Kuhl, M. Leydet, A. F. Lotter, P. E. Tarasov, and S. Tonkov. 2009. The European Pollen Database: past efforts and current activities. Vegetation History and Archaeobotany 18:417-424.*
>
> *Giesecke, T., B. Davis, S. Brewer, W. Finsinger, S. Wolters, M. Blaauw, J.-L. Beaulieu, H. Binney, R. Fyfe, M.-J. Gaillard, G. Gil-Romera, W. O. Knaap, P. Kuneš, N. Kühl, J. N. Leeuwen, M. Leydet, A. Lotter, E. Ortu, M. Semmler, and R. W. Bradshaw. 2014. Towards mapping the late Quaternary vegetation change of Europe. Vegetation History and Archaeobotany 23:75-86.*

Acknowledgment

Please use the following or similar acknowledgment: *“Data were obtained from the Neotoma Paleoecology Database (http://www.neotomadb.org, http://doi.org/10.17616/R3PD38), and the work of the data contributors, data stewards, and the Neotoma and in particular the EPD/ALPADABA community is gratefully acknowledged.”*