Best Practice Guidelines for Repositing and Disseminating Contextual Data Associated with Vertebrate Fossils

Released December 27, 2016

I. Introduction

The Society of Vertebrate Paleontology (SVP) strives to promote reproducibility of research results by ensuring that scientifically important vertebrate fossils and their contextual data are placed permanently in public-trust repositories to make them always accessible for researchers seeking to verify past results and to conduct new studies. By recognizing that fossils and fossil localities are nonrenewable resources, SVP strongly supports legal regulations such as those provided for the Paleontological Resources Preservation Act (PRPA; 16 U.S.C. §§ 470aaa – aaa-11) of 2009 that call for management and preservation of paleontological resources based on scientific principles and drawing on scientific expertise.

SVP’s Code of Ethics makes it the responsibility of vertebrate paleontologists to assist government agencies in developing management policies and to comply with those policies once enacted (Article 12, Section 2 of SVP Bylaws). In the U.S., PRPA Section 6304(c)(3) requires that the location of fossils collected on Federal land “will not be released by the permittee or repository without the written permission of the Secretary.” The passage of PRPA creates an opportunity for our Society to review best practices for distribution of contextual data (i.e., locality and stratigraphic information) for all vertebrate fossils, not just those from U.S. Federal lands.

This document describes professional best practices for repositing and disseminating contextual data associated with vertebrate fossils, especially locality and stratigraphic information. A working group led by SVP’s Government Affairs Committee assembled this guide with input from field paleontologists, museum researchers, collection managers, and database managers. By following these best practices, researchers and repository personnel will facilitate preservation of paleontological resources in the field and maximize the scientific value of research collections now and in the future. These recommendations draw on recommendations from the Global Biodiversity Information Facility (GBIF) Guide to Best Practices for Generalising Sensitive Species Occurrence Data (http://www.gbif.org/resource/80512). Although the best practices presented here are meant as guidelines, SVP expects members to be aware of these professional standards as they develop and to provide the Society with feedback to help improve them.

II. Principles

As required by SVP’s Code of Ethics, pertinent contextual data, including location and stratigraphic information, should be accurately recorded for every vertebrate fossil at the time of collection. These data should be archived along with the specimens in an appropriate, publically accessible and permanent repository (Article 12, Sections 1 and 4 of Bylaws). The contextual data augment the scientific value of the
specimens. In cases where release of these data will not create risk of harm to, or theft or destruction of, resources remaining at the collecting site (sometimes including non-paleontological resources), they should be disseminated freely to facilitate research, education, resource management, and other public benefit uses.

A. Contextual data must be archived with each specimen in a public-trust repository whose mission is to make both specimens and contextual data permanently accessible for scientific research, education, and other uses. Fossil locality data should be recorded to the greatest accuracy possible and fully georeferenced, down to meters or even centimeters where possible, and should be reposited along with stratigraphic and other metadata, detailed maps, measured sections, field notes, photographs, and other associated documents. The data should be precise enough that future researchers will be able to return to the collecting site when additional information or specimens are required. For fossils collected from regulated properties such as U.S. Federal lands, the contextual data should also be reported to, and archived with, each relevant land management agency.

B. Research publications that use contextual data should clearly identify the repository that houses the associated specimens and should explicitly reference the corresponding specimen and site catalog numbers so that interested parties can locate the original specimens and their associated data (i.e., reproducible published scientific results).

C. Wherever possible, contextual data should be disseminated widely and freely.

D. In some cases, public access to contextual data, especially the precise location of the collecting site, can result in harm to fossils, contextual information (e.g., taphonomic and sedimentologic data), or to non-paleontological resources (e.g., endangered species or delicate ecosystems) that remain in the field. In such cases, distribution of information may need to be controlled although the presumption remains in favor of release. Any restrictions placed on the dissemination of contextual data should be adhered to rigorously by the collector, repository, and all parties with whom the data have been shared.

E. The sensitivity of all contextual data, especially the location of the collecting site (criteria defined below), should be reviewed by an agency responsible for, or the owner of, the land where the fossils were collected, the repository, and the collector/permittee. In order not to hinder research, curation, and education, the review should be completed as expeditiously as possible.

F. Dissemination of contextual data should be restricted only when there is a genuine risk. Restricting contextual data may affect the precision of research based on aggregated data, such as analysis of fossil occurrences in the Paleobiology Database, iDigBio, or other such public data portals. Therefore, restrictions should be imposed only if absolutely necessary, whereas all contextual data should be made available for research upon request.

G. Repository managers and other data providers should consider the needs of users for access to contextual data and other documentation when they evaluate sensitivity and weigh the impacts of disseminating data and restricting their access. For paleontological sites on U.S. Federal lands that fall
under PRPA, this determination is, by law, the responsibility of the agency (permitter) that manages the land.

H. In cases where restrictions are placed on access to contextual data, the original data should be retained intact by the repository. Original data should never be altered, falsified, or discarded.

I. Because research depends on the accuracy of data, users should be informed about omissions or changes that have been made to metadata in the interest of protecting a site. In cases where redacted data are disseminated, especially cases where the precision of geographic coordinates or stratigraphic placement has been purposefully reduced to protect the location of the collection site, the fact that this has been done should be distributed as part of the metadata for that specimen. In public databases, such as repository catalogs or data aggregators like the Paleobiology Database, redacted records should be indicated with appropriate wording, rather than by leaving fields blank or null.

J. Users of contextual data should respect any and all restrictions that have been imposed by data providers. If granted enhanced access to restricted information, users must not compromise or otherwise infringe its confidentiality.

K. Whenever a data provider receives an application for access to restricted data, the assumption of continued sensitivity should be avoided. Rather, the occasion should be used as an opportunity to re-evaluate the determination. Decisions made by government agencies to release previously restricted contextual data must be made in consultation with the repository in order to meet the needs of non-Federal partners, the scientific community, and the general public. Cooperation with relevant governmental bodies is particularly important for repositories or situations where a ‘freedom of information access’ law applies in order to discuss potential ramifications of sharing requested sensitive information prior to its formal release.

III. Determining Sensitivity

Custodians of contextual data, including land managers, repository curators, repository collections staff, and data aggregators, are responsible for evaluating whether data in their possession should be regarded as ‘sensitive.’ Sensitive information is here defined as any contextual data, which, if released to the public, would create risk of harm to, or theft or destruction of, the resource or its origin (e.g., localities containing the paleontological resource), or adverse effect to a person or people. For vertebrate fossils, such potential harm most often stems from the possibility of damage to, or theft of, additional specimens of a rare taxon or additional parts of an individual organism that has previously been partially collected. Factors that should be taken into account when determining sensitivity include the type and level of threat, uniqueness of the taxon or attribute, type of information, and whether it is already publicly available. Note that non-paleontological factors may also make a site sensitive, including environmental and ecological factors. A criterion-based approach is outlined below to assist in evaluating these factors.
A. Criteria for Impact of Harm

An answer of ‘yes’ to any of the following criteria indicates potential of harm from human activity. The more ‘yes’ answers, the more likely the impact of harm there is.

1. Are there taxa that occur only at this site?
2. Do rare taxa occur at this site?
3. Does the site yield unusually complete or exceptionally well preserved fossils?
4. Is this site a unique or rare exemplar of its geological age or geographic provenance?
5. Is the site likely produce additional specimens that will add significantly to the knowledge of variation within a taxon or to the completeness of a paleocommunity?
6. Are contextual field data that remain at the site easily damaged or destroyed?
7. Have excavation and contextual documentation of the site been completed?
8. Are scientifically important fossils that remain at the site easily damaged?
9. Is the site easily located because of proximity to roads or to easily visited or identifiable landmarks?

B. Criteria for Determining Sensitivity

An answer of ‘yes’ to any of the following criteria indicates increased sensitivity of the data. The more ‘yes’ answers, the more likely the impact of harm there is.

1. Is the content and detail of the contextual data such that their release would enable someone to carry out activities that are harmful to fossils or data that remain at the site?
2. Would release of the contextual data bring them into the public domain for the first time?
3. Would release of contextual data damage a partnership or relationship that is essential for preserving the site?
4. Would increased traffic to the site endanger non-paleontological resources, including endangered species, environmentally sensitive areas, or native American cultural patrimony?
5. In cases where the contextual data do not include the precise location of a site, would its disclosure allow the location to be easily inferred from other publicly available information sources?
C. Recommendations for Georeferencing Based on Sensitivity Category

The above criteria should be used to assign a sensitivity category for sites using the following table. These criteria are intended to assist assessment, and are not prescriptive. Decision-making should incorporate the best available scientific principles and expertise. Geographic data should be disseminated with the following precision.

<table>
<thead>
<tr>
<th>Category</th>
<th>Sensitivity</th>
<th>Georeference Guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Extreme</td>
<td>Location not released, or released only as broad categorical information (formation/region/county, etc.) without georeference coordinates.</td>
</tr>
<tr>
<td>3</td>
<td>High</td>
<td>Georeference rounded to 0.01 degree (or nearest 1.5 km)</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
<td>Georeference rounded to 0.001 degree (or nearest 0.5 km)</td>
</tr>
<tr>
<td>1</td>
<td>Low</td>
<td>Georeference unrestricted (i.e., full precision released)</td>
</tr>
</tbody>
</table>

As a point of reference, 'high sensitivity' (i.e., two decimal places in decimal degrees) roughly correspond to resolution no greater detail than the nearest kilometer, or the last three places in Universal Transverse Mercator (UTM) coordinates.

IV. Implementation for Data Providers, Aggregators, and Distributors

Data custodians such as repositories or federal, state, and local governmental units should routinely provide complete unrestricted access to contextual data in their care unless these data have been assigned a sensitivity category of '2' or higher, as outlined above. Georeference coordinates for fossil occurrences should only be completely withheld for the most extremely sensitive sites. The following additional guidelines for dissemination of data should be followed.

A. Authors should provide georeferenced coordinates for fossil occurrences, preferably with maps and photographs, based on the sensitivity category of the site in question. For fossils with 'high' to 'extreme' sensitivity, publically accessible media (e.g., publications, public and conference presentations, social media, broadcast media, etc.) should not include images (photographs or video) containing landmarks that allow the sited to be easily located (e.g., highways, named/mapped geographic features, etc.). Authors should state whether the coordinates they provide are at full or reduced precision.

B. Data aggregators and distributors should continue to allow access to fossil site location data that are already in the public domain through scientific publications or occurrence databases and should add previously restricted occurrence coordinates whenever their sensitivity is deemed to have decreased to an appropriate level. Data aggregators must indicate via metadata those coordinates whose precision has been reduced when they redistribute georeference data.