

March 6, 2020

**Subject:** RFC Response: Desirable Repository Characteristics

Dear U.S. Office of Science and Technology Policy,

We represent the Society of Vertebrate Paleontology (**SVP**: <http://vertpaleo.org/>), a non-profit international scientific organization with over 2,000 researchers, educators, students, and enthusiasts. Our mission is to advance the science of vertebrate palaeontology (a discipline within life sciences) and to support and encourage the discovery, preservation, and protection of vertebrate fossils, fossil sites, and their geological and paleontological contexts. This letter is in response to the White House Office of Science and Technology Policy's (OSTP) for public comment on *Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting from Federally Funded Research* (85 FR 3083; pages 3085–3087; document number 2020-00689). All of our comments concern the middle and right columns on page 3086 85 FR 3083, including “I. Desirable Characteristics for All Data Repositories.” SVP does not have any specific comments on “II. Additional Considerations for Repositories Storing Human Data (Even if De-Identified).”

### **Types of Paleontological Data and Metadata to be Managed by Repositories**

We understand that the “proposed characteristics are intended to be consistent with criteria that are increasingly used by non-Federal entities to certify data repositories, such as ISO16363 Standard for Trusted Digital Repositories and CoreTrustSeal Data Repositories Requirements, so that repositories with such certifications would generally exhibit these characteristics” (page 3086). In addition to the requirement that all digital data from federally funded research should be repositied, SVP suggests that the language of this regulation be expanded to include the physical fossils collected by federally funded research. This is because physical fossils are also a form of data in the field of paleontology besides all associated information and generated data stemming from them, hereafter collectively referred to ‘paleontological metadata.’ Paleontological metadata, include, but not limited to:

- hard copy data (e.g., maps; photographs; field notes, including qualitative and/or quantitative measurements used or taken by researchers; catalog cards; letters containing specimen data; scientific illustrations; publications);
- digital data (e.g., various types of databases, including those that record locality and stratigraphic information, taxonomic and specimen catalogs, measurements, as well as names of land owners, collectors, donors, and/or preparators of fossils; digital photographs; 2-D and 3-D digital scan data; GPS coordinate data; electronic scans of hard copy data; electronic communication containing specimen data; publications);
- replicas (copies of fossils, including molds and digital data to make casts; 3-D prints based on digital data); and

- 'data reserves' for possible future studies, including chemical and microscopic analyses (e.g., rocks and sediment samples; fragmentary fossils; associated fossils collected with primary fossils).

The characteristics of an appropriate repository needed for best practices in paleontology are those that provide long-term preservation and access of not only digital data but also physical fossils and any other forms of paleontological metadata. Because science is an endeavor to make new discoveries, the types of metadata listed above should not be considered comprehensive, where presently unforeseen new types of paleontological metadata may come about in the future that repositories should also accommodate their storage and dissemination. In addition, paleontological metadata to be repositied may even include information in the absence of actual collected fossils. Examples include locality and stratigraphic data of known paleontological sites that have not yet been scientifically explored. Digital data in paleontology include those that represent 'extractions' from physical fossils (e.g., digital scan data as well as field photographs and notes when fossils were surveyed or collected) and therefore are implied pointers to information that is subject to verification. It must be noted also that such information and databases, regardless of whether or not any actual fossil specimens have been collected, often implicitly contain hypotheses or other potential intellectual properties. In addition, restoration and reconstruction of fossils, including physical skeletal mounts, restored fossil elements, digitally reconstructed anatomical elements or skeletons, or even scientifically-based artwork of extinct organisms (including digital images) should also be considered as forms of paleontological metadata where they potentially represent testable hypotheses.

From SVP's perspective, desirable repository characteristics are those that can accommodate management of all types of physical fossils and paleontological metadata. For physical fossil specimen care as well as paleontological metadata storage and dissemination, a wide range of capabilities exists. Efforts should be made by agencies to assist where possible with the ultimate goal of bringing each up to consistent standards. For practical considerations, inadequacies should not exclude granting or maintenance of repository status, but rather additional support should be given to such repository agencies or institutions to help bring them to consistent standards.

We would also like to have a clarification. As noted above, 3-D digital scan data that capture the three-dimensional likeness of objects, such as paleontological (as well as biological and archaeological) specimens, can allow for the reproduction of precise replicas of these objects for scholarly or commercial uses. In cases where these objects are owned by the Federal Government (i.e., original specimens collected from federal lands), reproduction rights are controlled by the permit agreements under which they were collected, and associated federal regulations. How will replica production be restricted, if at all? The rules should allow replica production at least for scholarly and educational purposes.

### **Desirable Characteristics of Paleontological Repositories**

The principle reason for placing scientifically important fossils in a public repository is that vertebrate fossils are rare and often unique. Scientific practice demands that conclusions drawn from the fossils and associated paleontological metadata should be verifiable: i.e., scientists must be able to reexamine, re-measure, and reinterpret them, where such reexamination can happen decades or even centuries after the fact. Furthermore, technological advances, new scientific questions, and opportunities for synthetic research mean that new research often utilizes fossils and associated paleontological metadata that

were originally collected with other purposes in mind. These lines of reasoning mandate that scientifically important fossils be preserved along with their associated paleontological metadata for decades, centuries, and hopefully millennia. Optimal characteristics of suitable repositories include:

- a primary mission that encompasses the preservation of scientifically important fossil specimens and associated paleontological metadata;
- a non-profit organizational structure that is capable of weathering economic changes, political changes, and other changes of fortune
- a demonstrated commitment to preserving specimens and to managing associated metadata such as locality and contextual information (see U.S. Department of Interior's guidelines for federally approved repositories and SVP's *Best Practice Guidelines for Repositing and Disseminating Contextual Data Associated with Vertebrate Fossils* ([http://vertpaleo.org/GlobalPDFS/SVP-Paleo-Best-Practice-Guidelines-\(2nd-Ed\).aspx](http://vertpaleo.org/GlobalPDFS/SVP-Paleo-Best-Practice-Guidelines-(2nd-Ed).aspx));
- a commitment to hiring staff with advanced degrees or equivalent training in paleontological science, curation, and preservation;
- a well-considered policy for keeping fossil specimens and their associated paleontological metadata in the public trust should circumstances change such that the repository no longer able to care for them; and
- a primary mission that includes facilitating active research on the repository's fossil and associated paleontological metadata holdings.

Appropriate repositories therefore include publicly accessible, non-profit museums, universities, colleges, geological surveys, and government agencies whose funding does not hinge on the success of a single company, whose mission statement includes research or education, and whose policies include protocols for keeping material in the public trust if the institution can no longer care for it. Institutions that are set up as non-profit organizations largely independent of the original benefactors would most likely be recognized as credible repositories by peers in the field of vertebrate paleontology.

### **Access and Dissemination of Paleontological Data and Metadata by Repositories**

Reproducibility of paleontological research rests on the premise of permanency and accessibility of examined fossil specimens as well as paleontological metadata, including digital data, deposited in stable repositories under public trust. Because fossils are nonrenewable resources where every fossil specimen is unique, storage of and access to them, along with all associated metadata, must be done with care by repositories. The presumption is that all fossil specimens and paleontological metadata, including digital data, curated by repositories remain permanently stored and accessible to anyone who wishes to access them. However, in some cases, public access to physical fossils and/or paleontological metadata in repositories may need to be controlled, especially if it can result in harm to the fossils, to on-going research, or to the fossil localities. In particular, data pertaining to specific locations of fossil collecting sites must be regarded as 'sensitive' where the following two conditions should be met before placing them in maximally open access data repositories: 1) for fossils collected from U.S. public land, clearance to release the geographic coordinates must be obtained from the relevant secretary as required by the Paleontological Resources Preservation Act (PRPA); and 2) for all paleontological sites, the sensitivity standards outlined in SVP's *Best Practice Guidelines for Repositing and Disseminating*

*Contextual Data Associated with Vertebrate Fossils* ([http://vertpaleo.org/GlobalPDFS/SVP-Paleo-Best-Practice-Guidelines-\(2nd-Ed\).aspx](http://vertpaleo.org/GlobalPDFS/SVP-Paleo-Best-Practice-Guidelines-(2nd-Ed).aspx)) should be followed. In addition to the details about our sensitivity standards, the *Best Practice Guidelines* also provides information concerning the handling of paleontological metadata, including digital data. Much of the following paragraphs come from the document, where the phrase ‘contextual data’ is replaced with ‘paleontological metadata’ for the purpose of this comment letter.

Wherever possible, paleontological metadata stored in repositories, including unpublished forms, should be disseminated freely and widely. However, in some cases, public access to paleontological metadata, especially the precise location of the collecting site, can result in harm to fossils, contextual information (e.g., taphonomic or sedimentologic data), on-going research, 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 in compliance with relevant laws and regulations as well as professional ethical standards, although the presumption remains in favor of release. Any restrictions placed on the dissemination of paleontological metadata should be well justified and adhered to rigorously by the repository as well as the collector and all parties with whom the data have been shared.

The sensitivity of all paleontological metadata, especially the location of the collecting site, should be reviewed by the repository as well as by the permitter (i.e., governing body responsible for, or the owner of, the land where the fossils were collected) and the permittee (i.e., collector/researcher) to the best of their ability. In order not to hinder research, curation, and education, the review should be completed as expeditiously as possible. Dissemination of paleontological metadata should be restricted only when there is a genuine risk to the collecting site. Restricting paleontological metadata may affect the precision of research based on aggregated data, such as analysis of fossil occurrences in online public data portals. Therefore, restrictions should be imposed only if absolutely necessary, whereas all paleontological metadata should be made available for research upon request.

Repository managers should consider the needs of users for access to paleontological metadata 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 the PRPA, this determination is, by law, the responsibility of the agency (permitter) that manages the land. In cases where restrictions are placed on access to paleontological metadata, the original data should be retained intact by the repository, and original data should never be altered, falsified, or discarded. Because research depends on the accuracy of data, repositories should inform the data users 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 (e.g., online data portals), redacted records should be indicated with appropriate wording, rather than by leaving fields blank or null.

Whenever a repository 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 paleontological metadata must be made in consultation with the repository in order to meet the needs of non-governmental 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.

Repositories acting as data custodians are responsible for receiving, maintaining and preserving all paleontological metadata related to localities, specimens, and collection acquisitions. While these data are maintained in public trust, complete access to data may be restricted at the discretion of the data custodian or as required by law. In the event that data are restricted, the repository manager should disclose this fact to data providers as well as data aggregators and distributors, or should include descriptive language to this effect on their respective online search forms. Should the extent of publicly available paleontological metadata prove insufficient for a given purpose, data users are encouraged to contact individual repositories for more specific inquiries. Repository managers should assess the needs of the user and the fitness for use of the request. Besides their names and institutional affiliations, data users may be asked to provide the following justification to repository managers: 1) a description of the data they seek to obtain; 2) a description of their research, education, resource management, or other public benefit project, and why the requested data are pertinent or essential to their research questions; and 3) a description of how they intend to use and disseminate the data if the request is granted. Repository managers are responsible for relaying institutional policies and specifying any terms and conditions that may be placed on information for release. It should be noted that paleontological metadata are not necessarily always precise, accurate, complete, or reliable. Records may be unverified, vague, contain inherent errors, or reflect incorrect data. Data custodians should impress the importance of not using search results uncritically, as failing to acknowledge these limitations may undermine the legitimacy of certain data interpretations.

We have one question concerning the dissemination of paleontological metadata, including digital data. In the case of data that are exempt from Freedom of Information requests so as to protect in situ scientific (and cultural) resources, such as paleontological (and archaeological) site data, how will these data be protected, and what information would the Persistent Unique Identifier (PUID) or Digital Object Identifier (DOI) point to?

Thank you for the opportunity to comment on this very important issue for scientific advancement. Comments and questions concerning this comment letter and/or our *Best Practice Guidelines* can be addressed to any one of us (our e-mails given below) or Dr. Kenshu Shimada (Chair of SVP's Government Affairs Committee: [kshimada@depaul.edu](mailto:kshimada@depaul.edu)).

Sincerely yours,

Emily J. Rayfield, Ph.D.  
*SVP President*  
[e.rayfield@bristol.ac.uk](mailto:e.rayfield@bristol.ac.uk)

Jessica M. Theodor, Ph.D.  
*SVP Vice President*  
[jtheodor@ucalgary.ca](mailto:jtheodor@ucalgary.ca)

P. David Polly, Ph.D.  
*Past SVP President*  
[pdpolly@indiana.edu](mailto:pdpolly@indiana.edu)

