

DATA ACCESS AND MANAGEMENT PLAN

Data access is a fundamental component of this project. The data products include 1) 3D digital data of museum specimens; 2) algorithms for shape analysis; and 3) publications and conferences.

(1) 3D digital data: I fully recognize the importance of promoting and facilitating rapid data sharing and access, as well as the need for archival stability of that access. With this in mind I have developed a database in collaboration with Whirl-i-Gig, Inc. that is designed and devoted to storing and distributing microCT data, 3D digital models, and flat photographic images of museum specimens. The database, **MorphoSource.org**, is based upon Whirl-i-Gig's *CollectiveAccess*, (<http://www.collectiveaccess.org>) GPL v3 licensed open source collections management software for archives and museums in use at hundreds of sites around the world. It is currently live and accessible by any interested researcher. It already hosts 1,916 specimens for public consumption (more than triple of what was present at the previous submission). *I constructed this site because there was no existing site to which investigators could upload their 3D data sets, attach metadata in a standard way, and then immediately and interactively share the datasets with other researchers.* Cornell's arXiv, Dryad, and Figshare are not available or suitable for my goal of loading massive amounts of digital data and building a "digital museum collection." I opened communications directly with the administrators of Digimorph (<http://digimorph.org>) and learned that this website will not accept their datasets for archive purposes and cannot make them accessible for live download without additional funds and significant lead-time for further infrastructure development. Though Phenome10k—hosted by a researcher at UCL—came online this year and represents a concept similar to MorphoSource, it does not have the depth of infrastructure needed to support and coordinate long-term orderly growth, or usability by a large community of researchers. It also lacks a system for allowing museums to monitor, regulate, and proof datasets representing material from their collections. **It should be noted that MorphoSource is being developed as a community-wide solution**, and I have invited institutes with massive CT collections to disseminate their holdings through this database (including UT–Austin and Penn State). Dr. T. Ryan (see Letter of Collaboration) and I are working on plans to extend the digital infrastructure of MorphoSource to Penn State as the first step in creating a network of institutions



Getting Started

Find & Download Datasets

Useful Info

BROWSE

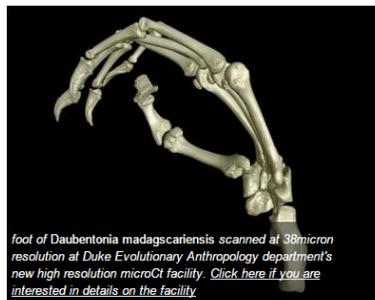
or

enter search terms

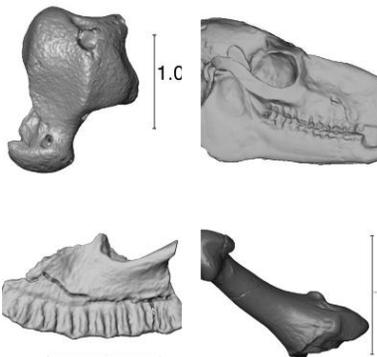


- [Information for Users](#)
- [Information for Contributors](#)
- [Terms](#)
- [User Guide](#)

LOGIN OR REGISTER



Recently Published



Welcome

MorphoSource is a project-based data archive that allows researchers to store and organize, share, and distribute their own 3d data. Furthermore any registered user can immediately search for and download 3d morphological data sets that have been made accessible through the consent of data authors.

The goal of MorphoSource is to provide rapid access to as many researchers as possible, large numbers of raw microCt data and surface meshes representing vouchered specimens.

File formats include tiff, dicom, stanford ply, and stl. The website is designed to be self explanatory and to assist you through the process of uploading media and associating it with meta data. If you are interested in using the site for your own data but have questions about security or anything else contact the site administrator. Otherwise please download whatever data you need and check back frequently to see what's new.

that support and manage this initiative.

MorphoSource is gaining broad support from museum curators and researchers given its structure and goals (see curator collaboration letters). It has been given strong endorsements on PLoS Blogs which reviewed the site in detail and concluded: "MorphoSource is functional and fairly easy to use. Currently, it's probably the best option for sharing raw CT data. Grant writers should feel comfortable listing MorphoSource in their Data Management Plan. Reviewers should feel comfortable asking authors post data to MorphoSource." (Werning, posted May 11, 2015)

MorphoSource is a web-based digital archive capable of accepting, preserving, and presenting a variety of data, including many image formats (e.g., TIFF, JPEG, PSD,

DICOM, and PDF), image “stacks” (sequences), video, documents, and 3D models (.ply, .stl, and .pdf).

MorphoSource is organized around projects created and managed by any registered and approved researcher. Each project has its own access control settings, allowing users to manage and coordinate public release of different data sets as they are completed. Once a project is published, its data are available to the public via text searches and a browsing system for online viewing and download.

I have negotiated an agreement with Duke’s Trinity Technology Service that will provide 10Tb of storage space for MorphoSource per year at no direct cost, with no limit put on the maximum size that MorphoSource can obtain and no limit on the duration for which it will be supported. By the time this grant is awarded and the project starts, 30Tb will be available. By the time the grant ends at least 80Tb will be available and this number will continue to grow. Such a commitment is exceedingly rare, if not unique, and its value cannot be overstated. This means that the service is free for use by independent, external researchers. In addition, Whirl-i-Gig will implement a web-services API enabling other databases to access and leverage MorphoSource-hosted data.

Timing of distribution. Datasets from this CAREER project will be loaded onto MorphoSource for eventual distribution as soon as they are generated. They will be published for public consumption 1) once an initial publication including them has been accepted; 2) no more than two years after their initial upload date; or 3) at my discretion (whichever comes first). External users of MorphoSource are not held to these requirements, but will be expected to eventually make their data accessible to others through the site, or begin contributing fees to help offset costs for the storage service. For the currently proposed project, I will include original TIFF stacks of scans, mesh files (which can be viewed in 3D and rotated through the web browser or downloaded and used for 3D printing), and 3D .pdf files for users with more limited access to 3D-viewing software. MorphoSource utilizes **creative commons** copyright permissions and licenses that allow third party use, but prohibit use for commercial gain.

An exciting promise of this initiative for 3D data distribution is that if MorphoSource’s popularity among researchers and collection managers around the world continues to increase over the next several years, it is entirely feasible that it will provision the “world’s collection” of particular species, available at researchers’ fingertips in a relatively short time. The importance of such a development for the global impact and democratization of science cannot be overstated. The bulk of most budgets contributed by NSF to evolutionary anthropology dissertations is for travel to museums to collect data on study samples. This also forms the most time-intensive part of many dissertations, such that time for analysis is often truncated. Partly as a result, such studies are primarily judged on the sample that was accrued. If the current project is funded, it will set the stage for an inversion of limiting steps to morphological research. Comprehensive digital archives like MorphoSource will allow time spent accessing research samples to be minimized and provide the potential for field-wide standardization of sample structure and size. This will allow more influential research to be done with smaller budgets, potentially decrease time for comparative morphology dissertations, and put a new emphasis on creativity and robustness of quantification protocols and analytical approaches.

(2) New Math and Analytical Functions: The PI and colleagues are implementing a workflow in which questions are posed and quantitative approaches are developed and refined by testing on large comparative data sets, and then finally coded in R, Python, and even more user-friendly Windows applications. These are distributed through CRAN, on my personal website, through MorphoSource, as well as in Cornell’s arXiv. The Matlab code for Boyer et al. (2011) is currently available through the Data Conservancy Project (arXiv #1110.3649) as well as at the website of the second author of this paper (<http://www.wisdom.weizmann.ac.il/~ylipman/CPsurfcomp/>). I have published R code for a second project (Boyer et al. 2015) and this is available <https://stat.duke.edu/~sayan/auto3dgm/index.shtml> as well as a MATLAB version that can be run on a cluster at <https://github.com/trgao10/PuenteAlignment>.

(3) Publication and Conferences: At least 10 major publications are planned as direct products of the proposed research. However, the actual number will probably be much higher as I publish around seven peer-reviewed papers per year directly related to either algorithm development or evolutionary history and functional morphology of Primates. Published works serve as critical validation for new algorithmic developments for the comparative method, and spread knowledge of the availability of specimen archives. There are plans and budget for attending multiple conferences each year: including American Association of Physical Anthropologists, Society of Integrative and Comparative Biology, and Society of Vertebrate Paleontology conferences. I will offer to speak on my work when I visit museum collections. Each year a workshop for disseminating information to and collaborating with educators is also planned.