

Data Management Plan

The *Pleiades* data management plan conforms to, and in many respects goes beyond the requirements of, New York University's *Policy on Retention of and Access to Research Data*, dated March 1, 2010.¹

Pleiades manages a variety of data classes that are necessary to support the description of ancient geography and to manage the activities of our scholarly community.² Upgrades planned for this project will refine the data model by adding support for relationships between places and associated documentation, but all existing content will be migrated with no loss of information. Data objects are versioned now, will continue to be versioned, and the versioning history of all data objects will also be migrated with no loss of information. Other datasets subject to this plan include blog posts, help documents, and other content published on the *Pleiades* web site, as well as conventional server access and error logs. *Pleiades* code is published on-line, and is also subject to this plan.³

In accordance with NYU policy, copies of all *Pleiades* research data will be retained by the PI for a minimum of 5 years following the final reporting of the project. Copies of some datasets will be retained for longer periods under various custodies (see below).

Server access and error logs will be retained in their native formats. Although they do not include personally identifiable information, they contain IP addresses and other sensitive information, and so will not be released publicly. Under terms of the NYU data retention policy, they will be retained and made available by the PI to appropriate individuals in the context of an audit, dispute, sponsor request, FOIA request, or other matter as directed by the Senior Vice Provost for Research (or a designee).

Pleiades geographic data is published over the web from the system in a number of standard formats, and we will maintain the full suite of these going forward. Individual data serializations for each *Pleiades* place, as described in the “Innovations” section, are disseminated using HTTP via stable URIs. The current nightly export to CSV format will continue to be disseminated via the *Pleiades Downloads* page at <http://pleiades.stoa.org/downloads/>. It will be enhanced to include all published data items and their attributes, thereby constituting a complete, application-neutral copy of the gazetteer in a widely used, non-proprietary open format that is easy for digital repositories and collaborating third parties to maintain and use. Textual information in all files is encoded using the standard UTF-8 (Universal Character Set Transformation Format 8-bit). This export package is currently accompanied by a ReadMe.txt file explaining the content and its expected use. The PI, in collaboration with the Managing Editors, will enhance this metadata with the addition of appropriate documentation in accordance with ISO 19115:2003 “Geographic information – Metadata (corrigendum 1)” as recommended by the Federal Geographic Data Committee (FGDC).

Pleiades is currently hosted on a leased server in a redundant data center located near Denver, CO. Our present contract provides for a dual, high-performance disk array configured as RAID 1, which provides

¹ NYU policy: <https://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/retention-of-and-access-to-research-data.html>.

² See further: *Pleiades Data Model*: <http://pleiades.stoa.org/help/pleiades-data-model>.

³ <http://pleiades.stoa.org/software/>.

for operational redundancy. The vendor provides comprehensive daily backups that are tested for restoration readiness on a quarterly basis (we see the raw test results on these). The contract also provides for 15-minute response time on-call emergency service and repair/restoration in the event of hard drive or other equipment failures. We believe that similar arrangements, adjusted as necessary over time to support performance of the web application and manage costs, are adequate to ensure data integrity and access to live *Pleiades* content.

In addition, we will implement a preservation strategy for *Pleiades*' geographic content by putting processes and agreements in place to effect the regular alienation and backup of the comprehensive *Pleiades* export in CSV format in multiple external hands. These provisions will ensure long-term access even in the event of a catastrophic event involving our datacenter, the demise of the project, or changed priorities on the part of New York University. The Ancient World Mapping Center at the University of North Carolina at Chapel Hill already copies the *Pleiades* CSV export to its own servers daily, and will continue to do so during and after the period of performance. The Perseus Digital Library at Tufts University has pledged to download and maintain copies in a similar manner, alongside its backups of its own data, beginning in September 2015.⁴ ISAW's digital programs team will continue to make quarterly deposits of the latest version of this file to NYU's Faculty Digital Archive (<http://archive.nyu.edu>), a formal, institutional archive operated by the NYU Library system that provides automated bit fixity testing and repair, geographically separated replication, and the assignment and maintenance of persistent identifiers for deposits. FDA deposits will be made with the addition of Dublin Core metadata conforming to NYU specifications.⁵ We will make parallel deposits of the CSV export and of all our software releases to the Zenodo.org research output repository, which is funded by the European Union and managed by CERN. Zenodo.org will preserve and assign a Digital Object Identifier (DOI) to each. The Managing Editors are actively cultivating additional, complementary partnerships, likely to include both informal holding of copies and formal archival deposit.⁶

As described in the "Significance" section, *Pleiades* assigns stable identifiers to all objects of interest in its dataset and serializes these identifiers into stable HTTP URIs. As a matter of policy, the *Pleiades* Managing Editors are committed to the persistence of these URIs with the support of ISAW, which budgets for the on-going registration of the *pleiades.stoa.org* domain. As a hedge against longer-term technological changes that might render the resolution of HTTP URIs and Internet domain names obsolete, *Pleiades* embeds the full URIs in every dissemination and archival format, thereby enabling future computational or manual reconstruction of connections to other datasets.

⁴ Crane letter.

⁵ FDA URIs conform to the *Handle Specification*: <http://www.handle.net/>. Dublin Core: <http://dublincore.org/>.

⁶ Discussions underway with the German Archaeological Institute and The Digital Archaeological Record (tDAR: <http://www.tdar.org/>).