

# Ancient Graffiti Project (AGP): Data Management Plan

## Expected data

---

We expect to produce several types of data:

- A graffiti-specific metadata schema
- A custom vocabulary for describing figural graffiti, published according to SKOS specifications
- An openly available database of graffiti in Herculaneum that adheres to the metadata schema and applies the controlled vocabulary
- An open RESTful API for searching for graffiti by all metadata fields
- The source code for a web application interface to search for graffiti using the search API and visualize the graffiti locations at multiple geo-spatial levels

## Period of data retention

---

The graffiti metadata will be published in, i.e., accessible from, the AGP and EDR and EAGLE web interfaces and APIs. As we complete our fieldwork, we will improve the entries (e.g., adding a graffiti's letter height). We favor partial entries over not publishing. By the end of the grant period, the metadata of the graffiti found in Herculaneum will be complete.

We will release the source code for each component of our platform after we have completed a version of the component. For example, we will release the code that drives the RESTful API when we have implemented and thoroughly tested searching by location. Again, we favor frequent releases with only partial functionality implemented rather than waiting to release. At the end of the grant period, we will release our fully functional code, with documentation on how to use the code.

## Data formats and dissemination

---

Our data will be disseminated in common formats and will adhere to best practices. The graffiti metadata schema will be published on AGP's web site. The graffiti metadata will be available from our open API in JSON as well as from the web interface. Since Cultural Heritage belongs to Italian State, EDR will host photographs of all inscriptions in accordance with EDR's memorandum of understanding with the Ministero per i Beni e le Attività Culturali (Ministry of Culture), accordo Mibac-EAGLE, dated 21 November 2005. The vocabulary to describe figural graffiti will be published according to SKOS specifications. The source code will be released under the Creative Commons license Attribution-NonCommercial-ShareAlike 2.5.

## Data storage and preservation of access

---

### AGP's Graffiti Metadata

Our graffiti metadata database is running on an Ubuntu virtual server. By using a virtual server, we can more easily upgrade the server's capabilities, thus providing better service to those accessing the data.

W&L's Information Technology Services (ITS) will provide resilient, high-speed storage, secure access, and regular backups to the graffiti data, which will be stored on mirrored systems in two data centers. The Richard A. Peterson Center, the primary data repository, is a state-of-the-art facility utilizing current best practice technologies to monitor and manage the equipment. Redundant HVAC, multiple power distribution units, uninterruptable power supply, automatic transfer switch, and generator provide a high

level of resiliency for the supported equipment and data. Regular 24-hour monitoring of equipment and services provided in this data center are also in place. The backup data center in Wilson Hall provides similar capabilities.

ITS uses Symantec Backup Exec 2010 to manage data backups using a direct to tape, monthly full backup/weekly incremental tape rotation. Tapes will be stored off site for archive purposes. Monthly tapes will be available for 12 months. After that period, an annual archive will be the primary restore point. Off-site storage of back-up data includes direct-to-disk, direct-to-tape and disk-to-tape duplications. Backup rotations vary depending on the data type and determined need for retention period, recoverability and change volatility. We will continually work with ITS to determine which backup rotation is most appropriate for the generated data. In addition to the backups at W&L, some of the graffiti data will be either housed (e.g., images) or duplicated (e.g., id number, find spot) on EDR's servers.

W&L's ITS uses VMware ESXi on HP blade servers to support the virtual server environment. Each virtual machine is able to run on any of eight different blade servers in the separate data centers. Storage systems in use are Tegile hybrid storage arrays. These systems use both solid state drives and spinning disks to provide high input/output operations per second. Each storage system uses dual/redundant controllers to provide access to the data. Daily backups of the virtual machines are run and stored to separate storage systems. Those backups are then moved to tape for longer term, off-site storage.

#### **AGP's Source Code**

The source code that drives the web interface and open API will be released periodically on GitHub: <https://github.com/AncientGraffitiProject/AGP> . In addition, we are using local Subversion repositories of the source code.