

Data Management Plan

Expected data

Type of data	When shared?	Under what conditions?
Gazetteer place name and locational dataset	At conclusion of start-up period, when the gazetteer is launched publicly. Thereafter, daily data dumps will be archived for free downloading in multiple formats (csv, kml, xml)	Free; sharing and remixing permitted under Creative Commons Attribution 3.0 License (CC-BY)
Tasa manuscript dataset (TEI XML; serialized to HTML5 and other formats [see below])	At conclusion of start-up period, when gazetteer is launched publicly.	Free; sharing and remixing permitted under Creative Commons Attribution 3.0 License (CC-BY)
Road network/LCP dataset (serialized to GeoJSON, PNG, TIFF, JPEG, KML/KMZ)	At conclusion of start-up period, after public launch of site	Free; granted by permission from Qhapaq Ñan Project, Ministry of Culture (Peru)
Open source code for LOGAR gazetteer	At conclusion of start-up period, after initial testing	Free; distributed via GitHub
Open source code for collaborative geospatial platform	At conclusion of start-up period, after initial testing	Free; distributed via GitHub
Progress reports (multimedia)	Ad hoc basis on project website	Free; sharing and remixing permitted under Creative Commons Attribution 3.0 License (CC-BY)
Report on gazetteer development	After the project has been completed	Free; sharing and remixing permitted under Creative Commons Attribution 3.0 License (CC-BY)
Report on geospatial platform development	After the project has been completed	Free; sharing and remixing permitted under Creative Commons Attribution 3.0 License (CC-BY)
Final report to NEH	After the project has been completed	Free; sharing and remixing permitted under Creative Commons Attribution 3.0 License (CC-BY)

Period of data retention

Data will be retained on a project-specific server for a minimum of three years beyond the completion of the startup phase, and on other storage media in the Spatial Analysis Research Laboratory at Vanderbilt University for a minimum of five years beyond the completion of the start-up phase.

Data formats and dissemination

Open source software from the project will be freely available on GitHub to maximize accessibility. LOGAR data will be stored digitally as sing XML (eXtensible Markup Language), an international standard format for digital information. The project will also adopt a specific scholarly version of XML, the guidelines of the Text Encoding Initiative (TEI). The TEI guidelines (www.tei-c.org) provide a robust vocabulary of over 500 descriptive XML elements well-suited for storing historical data. The TEI guidelines have now been used for over two decades to encode information from hundreds of thousands of historical texts (for example <http://www.fihrist.org.uk/>). The TEI community is also one of

the largest scholarly organizations supporting humanistic digital scholarship. GeoPACHA, built on OpenGeo, uses open standards throughout its application stack, including OGC standards like WMS, WFS, WCS, and others. Data from both LOGAR and GeoPACHA will be serialized to several machine- and human-readable formats to ensure accessibility and reuse. For human usage, all TEI XML files are rendered into an HTML5 page accessible through an online app (built using eXist DB). In addition, data is also available in several export formats including MADS XML, KML XML, GeoJSON, RDF Triples, and Dublin Core metadata (in XHTML). Written reports will be available as html documents and/or pdf documents via the project website.

Data storage and preservation of access

Storing LOGAR records in TEI XML provides assurance that the project's data will be available for long term scholarly research. Storing GeoPACHA data using its open data standards assures long-term support. All data will be stored on a project-specific server physically located in the Spatial Analysis Research Laboratory, Vanderbilt University. The server will employ daily back up and will run on Red Hat Enterprise Linux. Web server applications will include the Apache HTTP Server with a proxy module for Tomcat. Security measures will include perimeter, datacenter, and host firewalls and a security module within Apache. Daily data backups will be performed to storage media located off site for physical security. Open source software will be stored on GitHub.