W07   Data Standards, Archiving, and Analytics in Forensic Anthropology

Franklin E. Damann, PhD*, DPAA CIL, 106 Peacekeeper Drive, Offutt AFB, NE 68113; Jeffrey James Lynch, MSc*, 501 Park Avenue, Apt 126, Omaha, NE 68105; Sachin Pawaskar, PhD*, University of Nebraska Omaha, 1110 S 67th Street, Omaha, NE 68182; Carl N. Stephan, PhD*, The University of Queensland, School of Biomedical Sciences, Saint Lucia, Queensland 4072, AUSTRALIA; Stephen D. Ousley, PhD*, Mercyhurst University, Dept of Anthropology/Archaeology, 501 E 38th Street, Erie, PA 16546; Nicholas P. Herrmann, PhD*, Texas State University, Dept of Anthropology, 266 ELA, 601 University Drive, San Marcos, TX 78666; Alexandria Frye, MA*, 1811 Medical Parkway, #836, San Marcos, TX 78666; Felix Engel, MA*, Albert Ludwigs University Freiburg, 29 Hebelstrasse, Freiburg, Baden-Württemberg 79104, GERMANY; Stefan Schlager, PhD*, and Albert Ludwigs University Freiburg, 29 Hebelstrasse, Freiburg 79104, GERMANY

After attending this presentation, attendees will understand the benefits of a unified data architecture and ontology of forensic anthropology data, which enables the development and implementation of software applications for data analytics. Attendees of this session will join a community of users and will gain access to open source software for recording and managing biometric data in forensic anthropology.

This presentation will impact the forensic science community by providing access to an ecosystem of software applications for forensic anthropology that facilitates casework analyses. Casework efficiencies are realized through a common ontology, enabling data sharing and opportunities for new methods. The ecosystem of applications is based on open source software that fosters collaboration and community engagement via appropriate interfaces and Application Programming Interfaces (APIs).

Disparate and fractured datasets preclude collaboration among various research groups. The purpose of this workshop is to promote a unified data architecture and ontology for forensic anthropology. In doing so, the consistency, quality, and usability of skeletal data improves, providing for more robust reporting and analytics capability, and enhances collaboration and efficiency among globally distributed forensic scientists, researchers, networks, and infrastructures.

This workshop begins with a series of presentations introducing the concepts of: (1) a unified data architecture and its application to forensic anthropology; (2) a well-defined data ontology; and, (3) an overview of data analytics and project pipelines. These concepts are then reinforced through tutorials and hands-on demonstrations.

The Resource Description Framework (RDF) is a standard for data integration, serving as a formal representation of knowledge and a conceptual data model in knowledge management systems. The RDF-based data standard (i.e., RDFBones) is presented with research data from the Forensic Anthropology Center Texas State (FACTS). The Commingled Remains and Analytics (CoRA) is an open source software application for recording and managing data from skeletal specimens. CoRA is a web-enabled relational database built on the open source PHP Laravel framework. It was created to manage multiple data types obtained from skeletal specimens in a large commingled assemblage. In doing so, data attributes (i.e., metric and non-metric data) are then searchable and filterable to propose linkages among specimens associated to single individuals. During these presentations, attendees will gain access to software, create user accounts, and realize the benefits by performing data input and aggregation into various information systems and run complex data queries across various data types.

The later presentations build upon the previously introduced concepts by demonstrating the benefits through analytical packages developed for specific tasks. A short hands-on introductory session to R programming will be provided to introduce the basics of computer programming to attendees and ensure users can draw on extensibility functions of R and hacking R program source code for their own research purposes. The analytical packages OsteoSort, TDStats, Skelet-o-matic, and recent updates to FORDISC® will be discussed.

OsteoSort is an osteological sorting package for R (www.osteosort.net) and provides tools to conduct pair-matching, articulation, and association analyses on large commingled assemblages. A related package, OsteoShiny, provides a graphical user interface to OsteoSort. Hands-on demonstrations will involve conducting analyses within R and through the graphical interface.

TDStats is an R application for standardizing and automating facial soft tissue thickness analysis for craniofacial identification and is freely available from the craniofacial identification resource hub CRANIOFACIALidentification.com. It is built under an exploratory data analysis framework and heavily utilizes plots rather than statistical significance tests to encourage comprehensive understanding of the data to draw scientific inferences and generate enhanced central tendency statistics for use in casework. TDStats uses base tcl/tk for graphical user interfaces.

Skelet-o-matic is a Microsoft® Excel® macro-enabled program that enables inventory of skeletal remains and the automatic generation of a colored skeletal homunculus. As an Excel® spreadsheet with set cells for data values, completed inventory forms can be easily data mined in R.

FORDISC® software is used to assess sex and ancestry and estimate stature, and upcoming versions will add modules for: (1) age estimation using transition analysis; (2) ancestry estimation using macromorphoscopic traits of the skull; and, (3) ancestry estimation using dental morphology traits. These modules will be available as free standalone programs.

Tying separate and related analytical packages together demonstrates the utility and efficiencies gained in analytical processes linked via a computing ecosystem. By creating a community of forensic anthropology users who contribute data to unified data architecture patterns, human skeletal variation among spatially and temporally disparate collections becomes attainable.

Information Management, Data Analytics, Forensic Anthropology

*Presenting Author