Dataset Review—*Seshat: Global History Databank*

Author/Compiler: Peter Turchin, Chair (University of Connecticut); Harvey Whitehouse (University of Oxford); Pieter François (University of Oxford); Kevin C. Feeney (Trinity College Dublin)

Repository: http://seshatdatabank.info/

Seshat is a massive database of global historical and archeological data to facilitate investigating socio-political organization of human societies and the evolution of civilizations. Initiated in 2011, the databank is the brain-child of a multidisciplinary group of academics (cliodynamicians, anthropologists, and historians) who saw the need for a digital resource that enabled researchers to rigorously test hypotheses regarding how complex historical phenomena unfolded. Seshat also performs a broader service by providing digital access to the combined knowledge of generations of historians and archeologists for social scientists and humanities researchers.

References above to "cliodynamics" and "hypothesis testing" may have already sent chills through the core of many historians among the audience, as variants of both topics served as sources of bitter arguments in the field for decades. Whichever side one may fall in the blood feuds over such topics, both researchers and students will find great value in engaging Seshat.

The primary mission of the databank is to assist social scientists in comparing rival quantitative hypotheses, using historical and archeological data. The data are global, covering a wide temporal scope that ranges from 4400 BCE to the present, depending on the geographic region of interest. Research assistants and postgraduate researchers collect the data under the supervision of a Regional Editor. The data are then vetted by expert historians and archeologists

prior to being added to the databank. Some 211,611 data points are currently coded into the databank, but only around 5% of this can be accessed by researchers based on geographic areas of interest and units of interest associated with a geographic region (such as polities in the region). There are also future plans to make the data available in machine-readable formats. The database is organized according to units such as Natural Geographic Areas (NGA), variables, and spatio-temporal coordinates. It will be useful to explain the meaning of units and variables within the context of Seshat.

The available data are categorized into 30 geographic regions outlined by NGAs, approximately 100 km2 block areas that cover regions of interest such as the Cambodian Basin, Upper Egypt, Iceland, Cuzco Province (Peru), Big Island Hawaii, and Middle Yellow River Valley. The, NGA comprise not only the basic spatial unit of the database, but also the basic category by which data are entered into the database. Examples of other units of the database are Polity (independent political units ranging from villages to empires); Religious System (similar to Polity, but instead reflecting religious authority); Interest Group (social groups that pursue some common interest). There are extensive coding procedures for these units to ensure consistency throughout the database. All the units except for the NGA can vary with the spatio-temporal coordinates—in other words, units such as Polity can vary in size with time.

Variables in Seshat facilitate the collection and categorization of socio-political and economic characteristics of human communities—in other words, data are collected for all the variables for every society included in the databank. These variables are organized under the following categories: Social Complexity Variables (e.g. Hierarchical Complexity, Bureaucracy Characteristics, and Law); Warfare Variables (e.g. Military Organization and Military Technologies); Ritual Variables (e.g. Largest scale collective ritual of the official cult); Institutional Variables (e.g. Limits on Power of the Chief Executive); Legal System (e.g. Property Rights); Equity (e.g. Discrimination); Social Mobility (e.g. Slavery); Religion (e.g. Religious Morality); Well-Being (e.g. Economic Well-Being, Biological Well-Being); Economy Variables (e.g. Energy Sector, Housing); Agriculture (e.g. Agricultural Land Use); Population (e.g. Estimated Carrying Capacity). The 79 variables are coded into the database through the units mentioned in the previous paragraph, so that information can be retrieved on the units of interest (such as geographic regions or polities.)

Although very thoroughly organized, the databank currently suffers from shortcomings due to the ambitious scale of the project. For instance, any historian specializing in the Ottoman Middle East who accesses the Upper Egyptian NGA for the spatio-temporal coordinate of 1683 CE – 1839 CE is in for an eye-popping experience. The general description of this period opens with the statement that the Ottoman Empire was at its most powerful. While this is a clear case of a

breakdown in the expert vetting process, and almost certainly not an isolated case, such missteps do not diminish the immense utility and potential of this project. They will be corrected in time, with increased participation of experts across academic disciplines.

A recent article authored by the directors and numerous contributors of the databank vividly demonstrated the potential of Seshat as a source for multidisciplinary research. The article established nine distinct characteristics that measured the developmental trajectories of 400 historical polities, describing the role of social complexity in the evolution of these polities. This is a precursor to tackling larger topics, such as a comparative analysis of different factors in the rise and evolution of human societies, including agriculture, warfare, trade, and political organization. Seshat is one of the few resources that offers paths for multidisciplinary research to approach such daunting yet fundamental topics.

Ahmet Izmirlioglu
Department of History
Utah State University
Izmirli73@hotmail.com