

## An Introduction to Seshat: Global History Databank

Please note: This is a preprint of a manuscript that has been accepted for publication in the *Journal of Cognitive Historiography*. The final published version may contain small differences such as formatting and copy editing. Please direct correspondence to [peter.turchin@uconn.edu](mailto:peter.turchin@uconn.edu).

Recommended citation: Turchin, P., Whitehouse, H., François, P., Hoyer, D., Alves, A., Baines, J., ... Xie, L. (In press). An introduction to Seshat: Global History Databank. *Journal of Cognitive Historiography*. doi: 10.1558/jch.39395

Turchin, Peter. University of Connecticut. [peter.turchin@uconn.edu](mailto:peter.turchin@uconn.edu)  
Whitehouse, Harvey. University of Oxford. [harvey.whitehouse@anthro.ox.ac.uk](mailto:harvey.whitehouse@anthro.ox.ac.uk)  
François, Pieter. University of Oxford. [pieter.francois@stb.ox.ac.uk](mailto:pieter.francois@stb.ox.ac.uk)  
Hoyer, Daniel. Evolution Institute. [dhoyer@evolution-institute.org](mailto:dhoyer@evolution-institute.org)  
Alves, Abel. Ball State University. [aalves@bsu.edu](mailto:aalves@bsu.edu)  
Baines, John. University of Oxford. [john.baines@orinst.ox.ac.uk](mailto:john.baines@orinst.ox.ac.uk)  
Baker, David. Macquarie University. [david.baker@mq.edu.au](mailto:david.baker@mq.edu.au)  
Bartkowiak, Marta. Adam Mickiewicz University in Poznań. [martabartkowiak86@wp.pl](mailto:martabartkowiak86@wp.pl)  
Bates, Jennifer. Brown University. [jennifer\\_bates@brown.edu](mailto:jennifer_bates@brown.edu)  
Bennett, James. University of Washington. [jsb11@uw.edu](mailto:jsb11@uw.edu)  
Bidmead, Julye. Chapman University. [bidmead@chapman.edu](mailto:bidmead@chapman.edu)  
Bol, Peter. Harvard University. [pkbol@fas.harvard.edu](mailto:pkbol@fas.harvard.edu)  
Ceccarelli, Alessandro. University of Cambridge. [ac2045@cam.ac.uk](mailto:ac2045@cam.ac.uk)  
Christakis, Kostis. British School at Athens. [knossoscurator@bsa.ac.uk](mailto:knossoscurator@bsa.ac.uk)  
Christian, David. Macquarie University. [director.bighistory@mq.edu.au](mailto:director.bighistory@mq.edu.au)  
Covey, Alan. University of Texas. [r.alan.covey@austin.utexas.edu](mailto:r.alan.covey@austin.utexas.edu)  
De Angelis, Franco. University of British Columbia. [franco.de\\_angelis@ubc.ca](mailto:franco.de_angelis@ubc.ca)  
Earle, Timothy K. Northwestern University. [tke299@northwestern.edu](mailto:tke299@northwestern.edu)  
Edwards, Neil R. Open University. [neil.edwards@open.ac.uk](mailto:neil.edwards@open.ac.uk)  
Feinman, Gary. Field Museum. [gfeinman@fieldmuseum.org](mailto:gfeinman@fieldmuseum.org)  
Grohmann, Stephanie. University of Edinburgh. [steph.grohmann@ed.ac.uk](mailto:steph.grohmann@ed.ac.uk)  
Holden, Philip B. Open University. [philip.holden@open.ac.uk](mailto:philip.holden@open.ac.uk)  
Júlíusson, Árni. University of Iceland. [arnidan@akademia.is](mailto:arnidan@akademia.is)  
Korotayev, Andrey. National Research University Higher School of Economics. [akorotayev@gmail.com](mailto:akorotayev@gmail.com)  
Kradin, Nicolay. Russian Academy of Sciences. [kradin@mail.ru](mailto:kradin@mail.ru)  
Kristinsson, Axel. Independent Scholar. [axel@akademia.is](mailto:axel@akademia.is)  
Larson, Jennifer. Kent State University. [jl Larson@kent.edu](mailto:jl Larson@kent.edu)  
Litwin, Oren. George Mason University. [olitwin@masonlive.gmu.edu](mailto:olitwin@masonlive.gmu.edu)  
Mair, Victor. University of Pennsylvania. [vmair@sas.upenn.edu](mailto:vmair@sas.upenn.edu)  
Manning, Joseph G. Yale University. [joseph.manning@yale.edu](mailto:joseph.manning@yale.edu)  
Manning, Patrick. University of Pittsburgh. [pmanning@pitt.edu](mailto:pmanning@pitt.edu)  
Marciniak, Arkadiusz. Adam Mickiewicz University in Poznań. [arekmar@amu.edu.pl](mailto:arekmar@amu.edu.pl)  
McMahon, Gregory. University of New Hampshire. [Gregory.McMahon@unh.edu](mailto:Gregory.McMahon@unh.edu)  
Miksic, John. National University of Singapore. [seajnm@nus.edu.sg](mailto:seajnm@nus.edu.sg)  
Moreno Garcia, Juan Carlos. French National Centre for Scientific Research. [jcmorenogarcia@hotmail.com](mailto:jcmorenogarcia@hotmail.com)  
Morris, Ian. Stanford University. [imorris@stanford.edu](mailto:imorris@stanford.edu)  
Mostern, Ruth. University of Pittsburgh. [rmostern@pitt.edu](mailto:rmostern@pitt.edu)  
Mullins, Daniel. University College London. [d.mullins@bbk.ac.uk](mailto:d.mullins@bbk.ac.uk)

Oyebamiji, Oluwole. Lancaster University. [o.oyebamiji@lancaster.ac.uk](mailto:o.oyebamiji@lancaster.ac.uk)  
Peregrine, Peter. Lawrence University. [peter.n.peregrine@lawrence.edu](mailto:peter.n.peregrine@lawrence.edu)  
Petrie, Cameron. University of Cambridge. [cap59@cam.ac.uk](mailto:cap59@cam.ac.uk)  
Prieser-Kapeller, Johannes. Austrian Academy of Sciences. [Johannes.Prieser-Kapeller@oeaw.ac.at](mailto:Johannes.Prieser-Kapeller@oeaw.ac.at)  
Rudiak-Gould, Peter. Independent Scholar. [PeterRG@gmail.com](mailto:PeterRG@gmail.com)  
Sabloff, Paula. Santa Fe Institute. [psabloff@santafe.edu](mailto:psabloff@santafe.edu)  
Savage, Patrick. Keio University. [psavage@sfc.keio.ac.jp](mailto:psavage@sfc.keio.ac.jp)  
Spencer, Charles. American Museum of Natural History. [cspencer@amnh.org](mailto:cspencer@amnh.org)  
Stark, Miriam. University of Hawaii at Manoa. [miriams@hawaii.edu](mailto:miriams@hawaii.edu)  
ter Haar, Barend. University of Hamburg. [barend.ter.haar@uni-hamburg.de](mailto:barend.ter.haar@uni-hamburg.de)  
Thurner, Stefan. Complexity Science Hub Vienna. [thurner@csh.ac.at](mailto:thurner@csh.ac.at)  
Wallace, Vesna. University of California, Santa Barbara. [vwallace@religion.ucsb.edu](mailto:vwallace@religion.ucsb.edu)  
Witoszek, Nina. University of Oslo. [nina.witoszek@sum.uio.no](mailto:nina.witoszek@sum.uio.no)  
Xie, Liye. University of Toronto. [liye.xie@utoronto.ca](mailto:liye.xie@utoronto.ca)

*Seshat: Global History Databank*, established in 2011, was initiated by an ever-growing team of social scientists and humanities scholars to test theories about the evolution of complex societies (Francois et al. 2016; Turchin et al. 2015). Seshat reflects both what is known about global history (within certain practical constraints, discussed below) and *also* what is unknown, or poorly known. Seshat is a continuously growing dataset incorporating evolving interpretations, highlighting persisting controversies, and contextualizing enduring ambiguities. The quantitative data, suitable for statistical analysis, is buttressed by qualitative nuance embedded in descriptive paragraphs along with references to pertinent scholarship.

A key innovation of the Seshat project is that it not only enables researchers to identify *static* patterns in the diversity and commonalities observed among past societies, but also to investigate *dynamic* processes that may generate cultural change. Thus, Seshat aims to construct temporal sequences recording how societies evolve and to explore the cross-cultural commonalities in how these processes have unfolded—surveying political, economic, religious, and other cultural characteristics along with relationships to biophysical systems (Currie et al. 2015; Turchin 2018; Turchin et al. 2019). It is part of a broader trend to use comparative historical data to address rigorously questions in the social sciences (Smith et al. 2012).

The first article to utilize fully the Seshat data was published in 2018 (Turchin et al. 2018). The most recent paper (Whitehouse et al. 2019) found that moralizing gods are not a prerequisite for the evolution of social complexity. Instead, beliefs in both powerful moralizing “Big Gods” and supernatural moralistic punishment more generally tend to appear after, rather than before, large increases in social complexity (*contra* the “Big Gods” hypothesis; cf. Norenzayan 2013; Norenzayan et al. 2016). Time-resolved data in Seshat, thus, were key in testing this hypothesis.

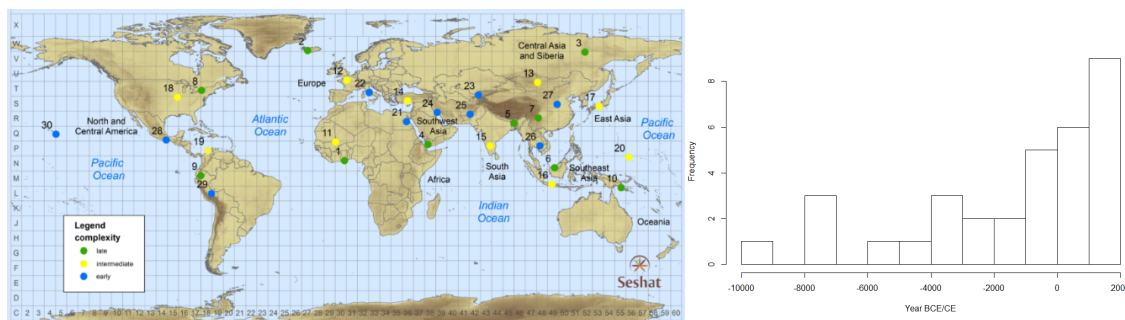
The possibility to adjudicate between different theories holds the promise of reshaping both social science and humanities research. Seshat Databank can also contribute to consilience between the two. Seshat provides a web of complex records connected along temporal, spatial, and thematic dimensions. Scholars have already started using this resource to compare characteristics among the societies they study and to trace the development of various attributes over time (Shin et al. 2019). Seshat can also be used to test theories about the co-evolution (or not) of social scale and complexity, agriculture, warfare, religion, and any number of such Big Questions. Seshat is rapidly becoming a massive resource for innovative cross-cultural and cross-disciplinary research.

## Temporal and Geographic Scope

Currently, Seshat focuses on the period between the Neolithic and Industrial Revolutions. Our unit of analysis is a *polity*, which we define pragmatically as an independent political unit ranging in scale from

villages (independent local communities) through simple and complex chiefdoms to states and empires. For each polity we code variables on social complexity, warfare, religion and rituals, agriculture, institutions, well-being, the production of public goods, and various technologies. The current [codebook](#) includes over 1500 variables, of which c.500 are the focus of active coding. As of May 2019, the Databank contains nearly 300,000 records linked to over 400 polities.

To address potential problems of selection bias and non-independence of data points, we use a [sampling scheme](#) based on 30 “Natural Geographic Areas” (or NGAs) across the globe. NGAs are simply spatial units that enable us to sample the diversity of past societies. First, ten major world macro-regions were identified (Figure 1a). To maximize potential variation, for each world region we selected one NGA in which social complexity—as defined in (Turchin et al. 2018)—emerged early, another where it emerged late, and a third in between. In different world regions, the time points at which social complexity emerged may be very different. For example, complex societies in Susiana, the early complexity NGA in Southwest Asia, go much further back in time than in Hawaii, the early complexity NGA for Oceania-Australia. NGAs are not the units for which we collect data, rather they help us select polities that become part of the Seshat sample. This approach yields a stratified sample of polities driven by geography and antiquity of social complexity, allowing analysts to construct spatially anchored time-series, while recognizing that the spatial extent of sampled societies fluctuates with time (as polities rise, expand, go into decline, and disappear).



**Figure 1.** (a) Locations of NGAs sampling global variation in cultural evolution. (b) Frequency distribution of the starting dates for data sequences in Seshat Databank. Adapted from (Turchin 2018).

## Data Collection

To populate the Databank, we consult current scholarship on each region along with expert collaborators to develop a list of all polities that sequentially occupied each NGA. In cases where NGAs encompass numerous coexisting small-scale societies, we treat these as *quasi-polities*, which roughly correspond to ethnological “cultures” (Murdock 1967; Murdock and White 1969) or archaeological sub-traditions (Peregrine 2003). We gather information on each variable in our codebook for each polity or quasi-polity, offering as much specificity as the evidence allows; often, a variable value is associated with the entire temporal extent of the polity, but if more granular information is known, we capture changes within the polity’s duration as well.

Prior to inputting data on a topic, we develop a conceptual scheme through Seshat workshops. The goal is to create a quantitative variable (e.g., polity population) or multiple proxy variables capturing various aspects of a more complex characteristic (e.g. well-being). Seshat research assistants (RAs) then code several test cases in consultation with experts, continually refining the variables. RAs are trained and supervised by teams of advanced (postdoctoral or professorial) social scientists and historians.

Supervisors facilitate communication between RAs and our expert collaborators, oversee coding decisions, and ensure consistent application of our coding schemes. We focus on hiring RAs with advanced qualifications and aim to retain productive researchers for lengthy periods, often several years.

Once a coding scheme is operationalized to test theories, data collection begins. First, RAs search the most up-to-date and relevant scholarship (with expert guidance and direct supervision by Seshat's more senior researchers), sourcing both primary and secondary material, and enter preliminary data. Second, RAs compile lists of questions on values that cannot be coded unambiguously, or on which information in the published sources is lacking, and seek help from the experts on the polity. Finally, we ask experts to review the data to check coding decisions made by RAs and help us fill gaps. The coding process is never "complete," as Seshat data are constantly checked by various stakeholders, and new relevant data may appear, or novel insights may alter the understanding of known documents and material.

When experts disagree or ongoing debate continues in the literature, multiple alternatives are coded. Likewise, when quantities are ambiguous, coders record the likely ranges [min, max] found in specialist literature. Where evidence is lacking entirely, RAs record the value as "suspected unknown", which becomes a question to bring to our collaborators; only domain experts can verify a value as "unknown", referencing their unique understanding of the limits of empirical evidence for a given area or topic.

We refer to a coded value of a variable for a particular polity as a "Seshat record" with a complex internal structure. Seshat records also include a narrative description explaining the background to the code and contextualizing levels of uncertainty and disagreement, along with citation to reference sources. The description may specify a quantity or range, or indicate whether a feature is present, absent, inferred present, inferred absent, suspected unknown, or unknown. "Inferred" presence or absence indicates some degree of uncertainty. For example, if iron smelting has been attested both for the period preceding the one that is coded and for the subsequent period, this could suggest a code of "inferred present" although there is no direct evidence for it (assuming general socio-technological persistence during this period and that there is no indication that this technology was lost and then regained). Variables can also have temporal uncertainty. For example, if we know that iron smelting appeared in a particular polity at some point between 300 and 600 CE, we code the period between 300 and 600 CE as effectively "either absent or present" (which is different from "unknown").

### **The Evolution of Seshat**

Seshat's methods have evolved since its inception in 2011. Initially, all data were collected by expert collaborators, but we soon realized that asking busy colleagues to fill in hundreds of boxes was misuse of their expertise. Having established an effective coding scheme, we discovered that much information can be accurately entered by well-trained and supervised RAs from published scholarship, allowing us to deploy expertise more strategically to resolve difficult coding issues, locate elusive information and point out relevant literature, and confirm genuine gaps in the record.

Seshat is a massive, complex project, which evolves constantly. In a project as vast and multi-faceted as Seshat, there will inevitably be some practical constraints on obtaining accurate or representative values or codes for specific variables because, for example, a particular bit of information has been published in an obscure source, or information of which we are not yet aware changes the coded value. We cannot wait until this "cleaning" process is over—because it never is. Our approach, thus, is to address such problems as we discover them, gradually improving the Databank, while understanding that some disagreement will always exist and that lag between our recorded data and available evidence will endure. We aim to bring as many voices as possible to bear on the information

we collect in order to utilize the most relevant scholarship in our published work. Once an article using Seshat Data is published, we present that material on our [website](#) as open-source data to be reused by other researchers. Knowing that the process of improving our data is never ending, we encourage users to offer feedback, pointing out alternate readings of evidence or directing us to previously unknown information.

The suggestions and critiques of scholars are essential in this regard. We all benefit by bringing out these issues into the open; the systematic nature of Seshat helps concentrate these discussions and identify where there are gaps in knowledge, uncertainties, and disagreements. Seshat is designed to be iterative—evolutionary—in nature. We are constantly rechecking coded data to make sure that the information we have is the most up-to-date and that it reflects relevant ambiguities and nuance. We are continually expanding both the geographic scope of data and adding new variables to cover different aspects of past societies. This is entirely a collaborative process, as Seshat functions only with the open, constructive engagement of researchers from a host of disciplines.

## References

- Currie, T. E., Amy Bogaard, Rudolf Cesaretti, Neil R. Edwards, Pieter Francois, Philip Holden, Daniel Hoyer, Andrey Korotayev, J. G. Manning, Juan Carlos Moreno Garcia, Oluwole K. Oyebamiji, Cameron Petrie, Peter Turchin, Harvey Whitehouse, and Alice Williams. 2015. "Agricultural Productivity in Past Societies: Toward an Empirically Informed Model for Testing Cultural Evolutionary Hypotheses." *Cliodynamics: The Journal of Quantitative History and Cultural Evolution* 6(1):24–56. doi: <https://escholarship.org/uc/item/4h29270b>
- Francois, Pieter, J. G. Manning, Harvey Whitehouse, Rob Brennan, T. E. Currie, Kevin Feeney, and Peter Turchin. 2016. "A Macroscopic for Global History. Seshat Global History Databank: A Methodological Overview." *Digital Humanities Quarterly* 10(4). doi: <http://www.digitalhumanities.org/dhq/vol/10/4/000272/000272.html>
- Murdock, George P. 1967. "Ethnographic Atlas: A Summary." *Ethnology* 6(2):109–236.
- Murdock, George P. and Douglas R. White. 1969. "Standard Cross-Cultural Sample." *Ethnology* 8(4):329–69.
- Norenzayan, Ara. 2013. *Big Gods: How Religion Transformed Cooperation and Conflict*. Princeton: Princeton University Press.
- Norenzayan, Ara, Azim F. Shariff, Will M. Gervais, Aiyana K. Willard, Rita A. McNamara, Edward Slingerland, and Joseph Henrich. 2016. "The Cultural Evolution of Prosocial Religions." *Behavioral and Brain Sciences* 39. doi: 10.1017/S0140525X14001356
- Peregrine, Peter. 2003. "Atlas of Cultural Evolution." *World Cultures* 14(1).
- Shin, Jaeweon, Michael Holton Price, David Wolpert, Hajime Shima, Brendan Tracey, and Timothy A. Kohler. 2019. "Human Societies First Grow, Then Improve Their Information Processing, Then Grow Some More." Preprint: *SocArXiv*. <https://osf.io/ejfbm>
- Smith, Michael E., Gary M. Feinman, Robert D. Drennan, Timothy Earle, and Ian Morris. 2012. "Archaeology as a Social Science." *Proceedings of the National Academy of Sciences* 109(20):7617–21. doi: 10.1073/pnas.1201714109
- Turchin, Peter. 2018. "Fitting Dynamic Regression Models to Seshat Data." *Cliodynamics* 9(1):25–58. doi: 10.21237/C7clio9137696
- Turchin, Peter, Rob Brennan, T. E. Currie, Kevin Feeney, Pieter Francois, Daniel Hoyer, Joseph Manning, Arkadiusz Marciniak, Daniel Mullins, Alessio Palmisano, Peter Peregrine, Edward A. L. Turner, and Harvey Whitehouse. 2015. "Seshat: The Global History Databank." *Cliodynamics: The Journal of Quantitative History and Cultural Evolution* 6(1):77–107. doi: <https://doi.org/10.21237/C7clio6127917>

- Turchin, Peter, Thomas E. Currie, Christina Collins, Jill Levine, Oluwole Oyebamiji, Neil R. Edwards, Philip B. Holden, Daniel Hoyer, Kevin Feeney, Pieter Francois, and Harvey Whitehouse. 2019. "Crop Productivity Estimates for Past Societies in the World Sample-30 of Seshat: Global History Databank." Preprint: *SocArXiv*. <https://osf.io/jerza>
- Turchin, Peter, Thomas E. Currie, Harvey Whitehouse, Pieter François, Kevin Feeney, Daniel Mullins, Daniel Hoyer, Christina Collins, Stephanie Grohmann, Patrick Savage, Gavin Mendel-Gleason, Edward Turner, Agathe Dupeyron, Enrico Cioni, Jenny Reddish, Jill Levine, Greine Jordan, Eva Brandl, Alice Williams, Rudolf Cesaretti, Marta Krueger, Alessandro Ceccarelli, Joe Figliulo-Rosswurm, Po-Ju Tuan, Peter Peregrine, Arkadiusz Marciniak, Johannes Preiser-Kapeller, Nikolay Kradin, Andrey Korotayev, Alessio Palmisano, David Baker, Julye Bidmead, Peter Bol, David Christian, Connie Cook, Alan Covey, Gary Feinman, Árni Daníel Júlíusson, Axel Kristinsson, John Miksic, Ruth Mostern, Cameron Petrie, Peter Rudiak-Gould, Barend ter Haar, Vesna Wallace, Victor Mair, Liye Xie, John Baines, Elizabeth Bridges, Joseph Manning, Bruce Lockhart, Amy Bogaard, and Charles Spencer. 2018. "Quantitative Historical Analysis Uncovers a Single Dimension of Complexity That Structures Global Variation in Human Social Organization." *Proceedings of the National Academy of Sciences* 115(2):E144–51. doi: 10.1073/pnas.1708800115
- Whitehouse, Harvey, Pieter François, Patrick E. Savage, Thomas E. Currie, Kevin C. Feeney, Enrico Cioni, Rosalind Purcell, Robert M. Ross, Jennifer Larson, John Baines, Barend ter Haar, Alan Covey, and Peter Turchin. 2019. "Complex Societies Precede Moralizing Gods throughout World History." *Nature* 568(7751):226-229. doi: 10.1038/s41586-019-1043-4