UC Riverside
Cliodynamics

Title

Seshat: The Global History Databank

Permalink

Journal

Authors

Turchin, Peter
Brennan, Rob
Currie, Thomas
et al.

Publication Date

2015

DOI

Supplemental Material

Copyright Information

Seshat: The Global History Databank

Peter Turchin,[1,2] Rob Brennan,[3] Thomas E. Currie,[4] Kevin C. Feeney,[3] Pieter François,[5,6] Daniel Hoyer,[2] J. G. Manning,[7] Arkadiusz Marciniak,[8,9] Daniel Mullins,[2,6] Alessio Palmisano,[4] Peter Peregrine,[10,11] Edward A. L. Turner,[5] Harvey Whitehouse[6]

[1]*University of Connecticut*
[2]*The Evolution Institute*
[3]*Trinity College Dublin*
[4]*University of Exeter*
[5]*University of Hertfordshire*
[6]*University of Oxford*
[7]*Yale University*
[8]*Adam Mickiewicz University (Poznań)*
[9]*Flinders University (Adelaide)*
[10]*Lawrence University*
[11]*Santa Fe Institute*

Abstract

The vast amount of knowledge about past human societies has not been systematically organized and, therefore, remains inaccessible for empirically testing theories about cultural evolution and historical dynamics. For example, what evolutionary mechanisms were involved in the transition from the small-scale, uncentralized societies, in which humans lived 10,000 years ago, to the large-scale societies with an extensive division of labor, great differentials in wealth and power, and elaborate governance structures of today? Why do modern states sometimes fail to meet the basic needs of their populations? Why do economies decline, or fail to grow? In this article, we describe the structure and uses of a massive databank of historical and archaeological information, Seshat: The Global History Databank. The data that we are currently entering in Seshat will allow us and others to test theories explaining how modern societies evolved from ancestral ones, and why modern societies vary so much in their capacity to satisfy their members' basic human needs.

Corresponding author's e-mail: peter.turchin@uconn.edu

Introduction

In 1919, the American archaeologist and historian James Henry Breasted, who held the first chair in Egyptology and Oriental History in the United States (and, in 1928, became President of the American Historical Association), issued a call for action:

> Here, then, is a large and comprehensive task—the systematic collection of the facts from the monuments, from the written records, and from the physical habitat, and the organization of these facts into a great body of historical archives. The scattered fragments of man's story have never been brought together by anyone. Yet they must be brought together by some efficient organization and collected under one roof before the historian can draw out of them and reveal to modern man the story of his own career. The most important missing chapters in that story, the ones which will reveal to us the earliest transition from the savagery of the prehistoric hunter to the social and ethical development of the earliest civilized communities of our own cultural ancestors—these are the lost chapters of the human career which such a body of organized materials from the Near East will enable us to recover (Breasted 1919).

Today, almost a century since Breasted wrote these words, his grand vision remains unfulfilled. The vast amount of knowledge about past human societies, held collectively by thousands of historians and archaeologists, has not been systematically organized. This knowledge is scattered across heterogeneous databases, innumerable books, publications in academic journals, and reports in the 'grey' literature, as well as notes in the private archives of scholars. Much of it is not even written, but resides in the heads of various experts and is permanently lost when they pass away.

The store of knowledge about past societies is now much more immense than it was in 1919, but it remains inaccessible for answering Big Questions about us and our societies, such as the one formulated by Breasted. Translating his question into more modern terms, we might state it as follows.

For most of our evolutionary history, humans have lived in small-scale societies of nomadic foraging bands that were integrated by face-to-face interactions and lacked both central authorities and high levels of structural inequality (Fried 1967, Lee and DeVore 1968, Service 1975, Mullins et al. 2013). The first large-scale, complex societies, characterized by an extensive division of labor, great differentials in wealth and power, elaborate governance structures, and large urban centers, appeared roughly 5,000 years ago (Liverani 2006, Algaze 2008, Wilkinson et al. 2014). How this "major evolutionary transition" (Maynard Smith and Szathmáry 1995) occurred is one of the biggest questions of social evolution, and is a question for which we still do not have a widely accepted answer. However,

we do know that a first step occurred in the Near East during the Early Neolithic period (ca. 9700/9500 – 6400/6200 BC) with the emergence of sedentary farming communities living in larger settlements, supported by a stable economy with surplus production(Price and Feinman 1995, Bar-Yosef 2001, Zeder 2011).

During the last 10,000 years, large-scale, complex societies have gradually replaced small-scale, foraging societies. Despite their ubiquity, the ability of today's large-scale societies to construct viable states and nurture productive economies, varies enormously from country to country. Why do states sometimes fail to meet the basic needs of their populations? Why do economies decline, or fail to grow? In many ways, differences between present-day societies can be as large as differences between our foraging ancestors 15,000 years ago and us today.

In their search for explanations of what makes some societies succeed and others "fail," most economists and political scientists focus on the current conditions or the recent past. Yet modern societies did not suddenly appear 30 or even 100 years ago; they gradually evolved from pre-existing societies over many centuries and millennia. If we want to live in better—more peaceful, wealthy, and just—societies, we need to understand the major evolutionary transition (or transitions) that occurred in our past and why they resulted in such divergent outcomes in the present.

Fortunately, today, we are finally able to rise up to Breasted's challenge. Our ability to do so is based on the remarkable evolution of information technology that has taken place over the last several decades. Even more important than improvements of hardware are recent developments in "knowledge engineering" techniques—approaches that will enable us to convert the gigantic and unorganized multitude of facts into structured knowledge, with which we can finally perform comprehensive tests of the many theories in history and cultural evolution. This body of organized knowledge will enable us to see much more clearly, if not fully recover, the lost chapters of a human career, thus fulfilling Breasted's dream.

This paper describes our vision for accomplishing this goal. We will do so by building a massive databank of historical and archaeological information, which we call Seshat: the Global History Databank. As this paper was written (in the first half of 2015), Seshat was rapidly becoming much more than a vision. Thanks to the generous support from government funding agencies, private foundations, and individual supporters (see Acknowledgments), we have already started the job of building it.

The goal of this article, thus, is to describe the structure and uses of Seshat. We start with a theoretical background—a brief overview of various theories that have been advanced to explain the emergence of large-scale, complex societies over the last 10,000 years or so, and the changing levels of inequality in the long-term history of our species. The main question in this section is, how do we extract

predictions from theories? Next, we explain how data on past societies are collected and organized in Seshat. In the conclusion, we discuss where we are and where we are going.

From Theories to Predictions

The Evolution of Social Complexity

Processes involved in the evolution of complex societies are hotly contested, with various theories emphasizing such factors as population growth, warfare, information management, economic specialization, and long-distance trade (Sanderson 1999, Johnson and Earle 2000). Generally speaking, theories tend to come in several flavors (Carballo et al. 2014). One factor that is common to virtually all explanations is that truly large-scale, complex societies can be built only on the basis of reasonably productive agriculture (Currie et al. 2015). Beyond the general agreement that agriculture is a necessary condition, however, opinions divide.

‘Functionalist’ (or ‘voluntaristic’) explanations emphasize benefits of cooperation to all: buffering environmental risk, managing competition and efficiently allocating resources, producing public goods such as an irrigation system, and capturing returns to scale, for example, in economic production (Service 1975, Johnson and Earle 2000). In contrast, ‘conflict’ explanations focus on the dark side of large-scale sociality: class struggle and exploitation, and warfare and conquest. For example, anthropologists influenced by the ideas of Karl Marx and Friedrich Engels argue that the adoption of agriculture created the surplus that could be then appropriated by the elites (Hayden 1995). In a recent book, Flannery and Marcus (2012) describe how the adoption of agriculture and the institution of private property could enable a few to amass great wealth while the majority was forced into servitude. Alternative explanations rely on conquest (Oppenheimer 1975) or, in more subtle versions, on environmental and social circumscription (Carneiro 1970).

Most current theories attempt to avoid the extremes of pure functionalism or pure coercion. The need for both cooperation and coercion in social life was clear to Ibn Khaldun (1958). Michael Mann (1986: 146-155) attempted to fold both forces into one concept, "compulsory cooperation." Robert Carneiro, who originally proposed that the transition from autonomous villages to centralized multi-village polities (chiefdoms) was accomplished by outright conquest (Carneiro 1970), shifted in later articles to a more cooperative model. He proposed that multi-village polities could have started as alliances of villages with military leaders; those leaders were able to transform their temporary powers over allied villages into enduring, centralized political structures (Carneiro 1998: 36). Yet other theories, such as the systems approach and cooperation and collective action perspectives, are discussed in Carballo et al. (2014).

The theoretical framework that has informed our research[1] is Cultural Multilevel Selection (CMLS). CMLS offers an elegant and parsimonious solution to the problem of how to combine the functionalist and conflict elements. Cooperation and coercion are enjoined in a very special way: cooperation takes place among lower-level units (but is supplemented with punishment of free-riders), whereas conflict takes place between higher-level entities (Richerson and Boyd 1998, Wilson 2002, Bowles 2009, Turchin 2013).

This brief overview shows that we do not lack theories. What we have lacked so far is a process by which we can reject some theories in favor of others. Different theories, however, make very different predictions as to where, when, and under what circumstances we should see the rise of large-scale complex societies during the last 10,000 years across the globe. Because different theories postulate different causal factors responsible for the evolution of social complexity, an analysis of which potentially explanatory variables correlate best with rising social complexity has a direct bearing on the empirical adequacy of rival theories. Moreover, we can do better than employing a purely correlational approach. Because causes precede effects, different theories make divergent predictions about temporal sequences of events. Thus, we need a *dynamical* databank that allows us to trace how different characteristics of past societies changed with time and determine which variables' changes precede changes in other variables. Of course, such an approach will not work in situations where temporal changes cannot be resolved on a sufficiently fine temporal scale; however, the continuous improvement in our methodologies of studying the past gives us hope that there will be a growing ensemble of case studies in which temporal resolution is sufficient for the purpose of testing theories.

In the next section, we make this discussion more concrete by focusing on one particular aspect of the evolution of human societies: the interaction between social complexity, hierarchy, and inequality.

The Evolution of Hierarchy and Inequality in Human Societies

The dramatic increase in the scale of societies over the last 10,000 years was accompanied by many other changes. One particularly interesting pattern is the trajectory followed by human inequality during this period (Figure 1).

---

[1] It is important to note that, although various members of our research network have particular theories that they are interested in testing, our overarching goal is to test rival theories against each other. Thus, data that we gather in Seshat focus on *theoretically relevant* variables (variables invoked by various theories, see Building the Databank below), but overall, the Databank is *theory-neutral*, and all effort is made to ensure that no particular theory is privileged. In the long run, theoretical neutrality is enforced by the open-ended nature of this collective enterprise, which allows for alternative conceptualizations and re-coding of any variable by proponents of different theoretical frameworks.
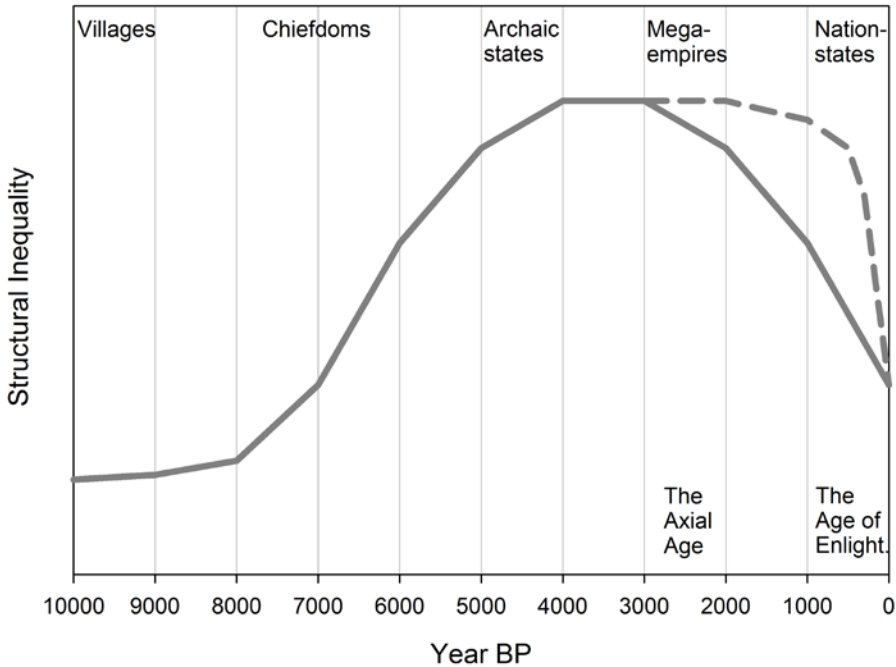
Figure 1. Two hypotheses about the evolution of human egalitarianism (with a focus on *structural* inequality). The main hypothesis that we plan to test proposes that structural inequality was low for small-scale polities, foraging bands and farming villages. It increased when first centralized polities, chiefdoms, evolved and reached a peak in archaic states. The Axial Age (800–200 BCE) was a turning point, and post-Axial polities, such as large empires and nation-states, were characterized by progressively lower structural inequality. An alternative to this 'Axial turn' hypothesis (the solid curve) is the 'Enlightenment turn' hypothesis (broken curve), positing that inequality remained at a high level until much more recently, the last two or three centuries.

Egalitarianism and a fierce preference for equality have characterized modern humans for the greater part of our evolutionary history. Although there are distinctions based on age, gender, and achievement, human foraging societies lack clear-cut dominance hierarchies, such as the ones present in chimpanzee troops. Equality in foraging societies is not simply a consequence of their relative poverty; it requires active maintenance. Egalitarian societies possess social norms and institutions designed to control those individuals who attempt to dominate others

and to obtain an unfair share of resources ('upstarts') (Cashdan 1980). These 'leveling mechanisms' range from gossip and ridicule to ostracism and, ultimately, assassination. Because of their small scale, societies of hunter-gatherers were integrated by face-to-face sociality, which enabled a diffuse, non-centralized form of social organization that was well-suited to maintaining an egalitarian ethos (Boehm 2001).

The adoption of agriculture c.10,000 years ago saw the evolution of larger-scale human societies (Flannery and Marcus 2012). More routinized forms of cooperation were required to sustain novel forms of specialized labor, reciprocity, pooling, and storage (Whitehouse 2004, Atkinson and Whitehouse 2010, Whitehouse and Hodder 2010). Agriculture, a sustainable exploitation of the commons, required the dissolution of small-group boundaries and inter-group rivalry in favor of larger-scale forms of collective identity, trust, and cooperation that extended to tens of thousands of individuals (Whitehouse and Lanman 2014), and, ultimately, millions. Computational demands on memory and information-processing systems increase dramatically with the size of the cooperating group (Dunbar and Shultz 2007). When compared with other organisms, humans have evolved to possess a number of cognitive advantages, including a more complex memory, more predictive capacities for future simulation, more fine-tuned systems for correctly identifying a number of individuals, and more precise numerical discriminatory abilities (Axelrod and Hamilton 1981, Fehr and Gächter 2002). As a result, humans are able to cooperate in the largest groups of all primates (Dunbar 1992). Despite these advantages, our cognitive systems are quickly overwhelmed when the size of a cooperating group grows beyond a few hundred.

For individuals to cooperate and coordinate their actions in larger groups, cultural workarounds are required to overcome these cognitive constraints (Richerson and Boyd 2001). The main solution that social evolution found was hierarchical organization, with large human groups integrated by chains of command. A member of a hierarchically organized group needs to have face-to-face interactions with only a few individuals: a superior and several subordinates (Turchin and Gavrilets 2009). This move to more hierarchical forms of social structure entailed formal offices of leadership and the development of hereditary systems of ranking and social class that were very different from the dominance hierarchies in non-human primates (Dubreuil 2008).

The advent of agriculture and hierarchical organization is intimately linked to a major sea change in the growth of structural inequalities. Although greater societal size would have had a number of advantages in terms of economies of scale and greater effectiveness in between-group competition, it required greater hierarchical complexity, resulting in the first turn in the evolution of egalitarianism (Turchin 2011). The nature and quantity of resources available to societies affects

how much wealth can be inherited, and whether inequalities can persist over multiple generations (Borgerhoff Mulder et al. 2009). Sedentary agricultural societies are able to produce food surpluses that can be used to support non-productive ('elite') members of society, whereas hunter-gatherer societies generally do not produce a surplus (Morrison 1994, Hayden 1995). The historical and archaeological records suggest that hierarchy and inequality increased after the advent of agriculture with the development of ranked societies and chiefdoms (Trigger 2003, Ames 2007). Sedentary farming may have served as a precondition for the rise of early states: a centralized political organization wielding its political, economic, and military authority over a territory and a defined group of people, guaranteeing the division of labor, the storage of food surplus, and the extraction of resources (Scheidel 2013). However, the most unequal—even despotic—human societies ever were 'archaic states' that first appeared c.3000 BCE (Feinman and Marcus 1998). These early states were characterized by extreme forms of structural inequality such as human sacrifice, slavery, unequal rights of commoners and nobles, and deification of the rulers (Trigger 2003, Kirch 2010). Religious and ritual mechanisms that evolved for the legitimation of hierarchy and structural inequality, initially serving the interests of society at large, may have been hijacked by coercive elites and rulers to drive inequality levels to unprecedented heights.

A major hypothesis, which we will test empirically, is that the Axial Age introduced another sea change in the evolution of inequality, starting a move towards greater egalitarianism that has been continuing to the present. The Axial Age (Jaspers 1953) refers to a series of religious and philosophical developments that occurred in such far-flung regions of Eurasia as Greece, the Near East, India, and East Asia between roughly 800 and 200 BCE. The last two centuries, have seen the spread of democratic forms of governance and widespread acceptance of fundamental human rights and equality. These are part of two related developments that may ultimately have begun in the Axial Age: (1) rulers have been increasingly constrained to act in ways that promote the public good, rather than their own interests, and (2) structural forms of human inequality have been gradually disappearing (the abolition of human sacrifice, slavery, etc.).

Religion may have played an extremely important, yet little-appreciated role in this second turn. Robert Bellah (2011) has recently argued that a major driver in the evolution of religion was the need to reconcile the tension between the benefits of hierarchy and the need for legitimacy and equity, resulting in the new forms of spirituality associated with the rise of world religions during the Axial Age. One aspect of this change was the first appearance of a universally egalitarian ethic, which was largely due to the emergence of "prophet-like figures who, at great peril to themselves, held the existing power structures to a moral standard that they clearly did not meet" (Bellah 2011).

The seeds of this transformation may be traced to pre-axial religions. Archaic states were nearly always characterized by some sort of divine kingship (Kirch 2010), and the evidence suggests that they practiced human sacrifice frequently and on a massive scale. While these are both indicators of extreme forms of inequality, during this phase, we also observe the appearance of 'gods,' who are distinguished from other powerful supernatural beings in that they are worshipped. Gods became an increasingly important part of social evolution during the Axial Age along with the rise of qualitatively new forms of social organization that employed new systems of legitimizing political power. During this time, gods evolved from capricious deities to transcendental and morally concerned supernatural beings (concerned, for example, with prosocial behavior by the rulers) (Norenzayan et al. 2015).

Furthermore, the very large-scale societies ('megaempires') of the Axial Age appear in many distant parts of Eurasia at approximately the same time. Geographically, these developments appear to be located in the region just south of the Great Eurasian Steppe. During the Axial Age, this vast region, stretching from Anatolia to North China, saw the recurrent development of large states and empires (Turchin 2009, Turchin et al. 2013). The new scale of larger empires, whose rulers had even more resources to aggrandize themselves, precipitated the legitimation crisis of the early axial state (Bellah 2011 following Jürgen Habermas). Axial religions may represent new systems of legitimizing political power, which were needed to prevent huge axial empires from splitting apart. The increase in size of the empires at this time may ultimately be linked to new forms of intense intersocietal competition.

From the perspective of CMLS, the evolution of cultural traits, such as religious practices and equity norms, can be modeled as resulting from the action of selective forces acting at different levels of social organization. Costly social institutions that enable large-scale cooperation can evolve and be maintained as a result of competition between societies: societies with traits that enable greater control and coordination of larger numbers will out-compete those that lack such traits (Richerson and Boyd 1998, Wilson 2002, Bowles 2009, Henrich et al. 2010).

Although societies can compete in many ways, one of the strongest and most important forms of competition is warfare, which may have acted as an important selective force favoring large-scale societies. The conditions of endemic warfare that are characteristic of small-scale societies (Keely 1997, LeBlanc 2003) may result in strong selection pressure for larger cooperating societies that can bring more warriors to a battle. Larger numbers alone do not guarantee victory, however, and military success is also aided by better coordination through centralized organization and chains of command. This may have been one reason why initially egalitarian human societies became more hierarchical and unequal (Turchin 2009, Bellah 2011).

However, while highly effective on the battlefield, a centralized military hierarchy has drawbacks as a general way of organizing societies—a society cannot really be held together by force alone. Additionally, great inequities resulting from rapacious military chiefs and their retinues alienate large segments of the population. As a result, early despotic chiefdoms and archaic states were very fragile and frequently did not outlast their founders. For example, oral records from Hawaii speak of popular uprisings against chiefs who extracted too much tribute (Kirch 2010). Bellah proposed that large-scale societies were only able to achieve stability by replacing brute-force domination by military chiefs with "a new form of authority, of legitimate hierarchy ... which involves a new relation between gods and humans, a new way of organizing society, one that finds a significant place for the disposition to nurture as well as the disposition to dominate" (2011). In other words, according to this theory, a major driver in the evolution of religion was the need to reconcile the tension between the need for hierarchy and the need for legitimacy and equity.

In summary, the evolution of social and economic inequality in human societies over the long term was not simply a U-turn from dominance hierarchies to egalitarianism and back to hierarchy, as suggested by some anthropologists (e.g. Boehm 2001, 2012, Flannery and Marcus 2012). Instead, there were two trend reversals: a 'zig' from small egalitarian groups to large-scale hierarchical and unequal societies, followed by a 'zag' from despotism towards the greater egalitarianism associated with the Axial Age. The hypotheses outlined above to explain these changes offer a set of predictions that differ in many respects from the predictions of rival theories and, therefore, can be tested with data on past societies. Table 1 lists these predictions together with possible alternatives.
It is important to note that Table 1 presents theoretical alternatives in a stark, binary fashion. In real life, of course, it is possible (indeed likely) that the best explanatory model will combine more than one mechanism, with different factors, perhaps, interacting in nonlinear, synergistic ways. Evaluation of such complex quantitative explanations is not a problem for modern methods of statistical analysis, especially when combined with a program of mathematical models that explicitly incorporate such interactions.

Building the Databank

Our long-term goal is for Seshat to be a vast repository for structured data on theoretically relevant variables for any past human society, for which such data exist. Our time frame is the last 10,000 years of human history (i.e., from the Neolithic to the present). 'Theoretically relevant' variables are those invoked by theories about how human societies evolve and function. Such variables either describe the phenomenon that we want to explain, or something that helps to explain the phenomenon: *explanandum* (for example, the trajectories of social

complexity, or of inequality) and *explanans*, respectively (Hempel and Oppenheim 1948). Some variables may play both roles, being interesting in their own right and also having an explanatory potential for other aspects of social dynamics. Because Seshat will eventually become an active databank, incorporating tools for automating the harvesting of variable values from Internet-accessible sources (e.g. historical and archaeological databases, DBPedia, JSTOR), the set of such variables will expand as new theories are proposed, invoking new explanatory factors. Seshat is a theory-neutral databank; future investigators will also have the freedom to re-conceptualize and redefine variables using their preferred theoretical frameworks.

A Focus on 'Big Questions'

In addition to having a long-term and highly ambitious goal, we need a set of realistic, medium-term objectives. Therefore, we initially focus on a limited set of variables. Our approach to selecting such variables is based on asking 'Big Questions,' such as the long-term trajectory of inequality, discussed in detail in a previous section (The Evolution of Hierarchy and Inequality in Human Societies).

Over the last several years, our research team has crafted and submitted a number of proposals to various funding agencies, each addressing a Big Question. Some of these proposals have been funded; others were declined. Reviewers' comments consistently indicated that the critical factor in success was the feasibility of the scientific approach and the potential societal impact of the Big Question under investigation. In effect, we have developed an 'evolutionary mechanism' that points us to the most interesting questions that can be feasibly addressed with our approach, as judged by the community of external expert reviewers employed by funding agencies (thus, the reviewers are the 'force of selection').[2]

It should be noted, however, that as Seshat grows by including more variables, it becomes easier to ask additional Big Questions. As we accumulate an increasingly large stock of *explanans* variables, we need fewer new variables to code. Furthermore, as we develop increasingly sophisticated software for automating the harvesting of structured data from semi-structured and unstructured sources, the effort required to capture the values of new variables will progressively decrease.

The World Sample-30

Just as our initial focus is on a limited set of variables, we similarly cannot start by trying to code all societies that have been studied by historians and archaeologists. This means we have to work with a sample. The need to sample societies, however,

---

[2] Successful proposals are listed in the *Acknowledgments*.

raises a number of conceptual issues. The defining characteristic of Seshat is that it focuses on the temporal development of variables. Thus, the potential of this approach to distinguish between rival theories critically depends on being able to follow evolutionary trajectories of societies through time. But how do we 'follow' societies?

Table 1. Examples of predictions generated by the Z-curve theory and possible alternative hypotheses.

Prediction	Explanation	Alternative
The Axial turn in structural inequality	Foragers and early farmers: Very low structural inequality Archaic: Very high struct. inequality Axial Age: turning point toward reduced structural inequality Post-axial: decreasing structural inequality	The turning point is modern: structural inequality only starts decreasing after 1700 CE, possibly associated with the Age of Enlightenment (18th century)
More-egalitarian societies (structurally or quantitatively) are more internally stable	Decreasing inequality helps in solving the coordination and cooperation problems present in large-scale societies by increasing the willingness of societies' members to solve such problems	Unequal societies are more stable. Rulers of more unequal societies can use the resources they extract from the population to better support their warrior retinues and bureaucratic administrators and impose stability by force
More-equal societies do better in between-group competition	More equal societies are more cohesive, better coordinated in military conflicts, and have individuals more willing to defend their group	More unequal societies can more effectively coerce larger numbers of people to fight in conflicts. Another alternative is that military effectiveness is relatively independent of broader societal cohesion (this may be particularly relevant to situations where societies have a permanent standing army)
Routinized rituals enabled the emergence of larger societies	Routinized rituals are necessary for the first appearance of large-scale, anonymous, hierarchical,	Large-scale, anonymous, hierarchical, centralized communities arose first (e.g., due to warfare), and

	centralized communities. Thus, they precede such large-scale societies	routinized rituals emerged subsequently to help maintain social cohesion
Axial religions made societies more equal by curtailing the coercive power of despots	The need to solve collective action problems entailed a return to more consensual forms of hierarchy, rather than forcing individuals to obey self-serving elites. Political and religious authority became more closely entwined, and leaders' legitimacy came to rely more on persuasion (e.g., through public, credibility-enhancing displays) and less on the naked exercise of power	Religions have been mainly used to support structural inequalities, such as the divine nature of rulers, legal distinctions between elites and commoners, or favor for one ethnicity over others. Ruling elites could use religion as 'opium for the masses,' which would legitimate the existing order (involving huge levels of inequality)
Intensified military competition leads to the appearance and spread of equity-promoting norms	More intense competition between groups favors more cohesive societies. Equity-promoting norms, including those hypothesized to stem from Axial religions, are a good way of achieving group cohesion	More intense forms of warfare select for more rigid, militarized hierarchies that concentrate power in the hands of rulers and elites, resulting in greater inequality
Warfare intensified during or just before the Axial Age	Intense competition between polities drove selection for cooperation-enhancing equity norms and institutions, resulting in nearly simultaneous appearances of Axial Religions in far-flung regions of Eurasia	Some other factor, such as global climate change, explains the simultaneous Axial Age developments in Southwest, South, and East Asia

One approach is to use geo-temporal sampling. First, we select a set of areas on the Earth's surface and a temporal sampling rate. We start with the approximate date at which agriculture was introduced to each area and advance through time to the present at intervals defined by the temporal sampling rate. At each point in time, we record the characteristics of the societies that occupy the area. A potential problem with this approach is that sometimes an entirely new group of people arrives from elsewhere and largely, or entirely, displaces the group that occupied the area previously. For example, until the seventeenth century, the Northeastern United States was occupied by a set of Native American societies. By the nineteenth

century, however, little trace of their social system remained. The European settlers brought their own set of genes, crops, domesticated animals, material culture, language, and institutions, which largely replaced the Native American equivalents. Thus, an alternative approach to geo-temporal sampling is 'ethno-temporal' sampling: following ethnic groups as they migrate and expand, shrink and die out, and recording the characteristics of the ethnic group at each point in time.

Neither approach is entirely satisfactory. Eventually, when our geographic coverage becomes more complete, we will be able to include the effects of both spatial proximity and 'cultural proximity' (for example, linguistic similarity as a proxy) in the analysis. Until then, however, we need to select one or the other. Because tracing ethnic roots of populations can be much more contentious, we opted for the geo-temporal approach. However, we emphasize that this is only the first phase in our long-term project. Eventually, our data should be complete enough to allow a simultaneous estimation of spatial and cultural proximity effects.

We selected 30 areas across the globe for our initial worldwide sample, stratified by world region and history of social complexity. We divided the world into ten major regions (see Figure 2 and Table 2) and then selected three natural geographic areas (NGAs, explained below) within each region. We looked for NGAs that sampled the diversity of a world region with respect to the relative antiquity of complex societies within it. Accordingly, one NGA was selected in an area that developed complex state-level societies very early. Another sampling point was the opposite in terms of the antiquity of complex societies—ideally, it was free of centralized polities (chiefdoms and states) until the colonial period. Finally, the third NGA was intermediate in social complexity. Because different world regions acquired complex societies at very different times, 'high complexity' NGAs can not be directly compared historically. For example, Susiana in Southwest Asia has much longer history of complex societies than Hawaii in the Pacific region.

In summary, the World Sample-30 was designed with two goals in mind: (1) to include as much variation amongst sampled societies as possible, at least along the social complexity dimension, and (2) to ensure that representation of different parts of the world was maximized.

Information Architecture of the Seshat Databank

The basic entity classes (types) in the databank's information architecture are Organizations and Territories, each of which has a set of temporally-scoped Variables associated with it. Entity classes serve as units of data collection and entry and as potential units of statistical analysis. Territories have fixed geographical bounds that do not change with time, whereas Organizations are defined by temporal bounds and may be associated with specific Territories at specific intervals. A subset of variables relate Organizations to Territories, and the

temporal dynamics of these *Relationship Variables* allow us to capture both temporal and geographical dynamics of the features of human societies. One of the most important relationship variables is the *controls* relation that specifies which social organization controls which territories over time. For example, a particular social organization known as a *polity* (defined below) may control a particular territory during a particular period of time, and several organizations, (e.g., interest groups, see below) may have the exists within relationship to a particular territory, which is controlled by a polity.
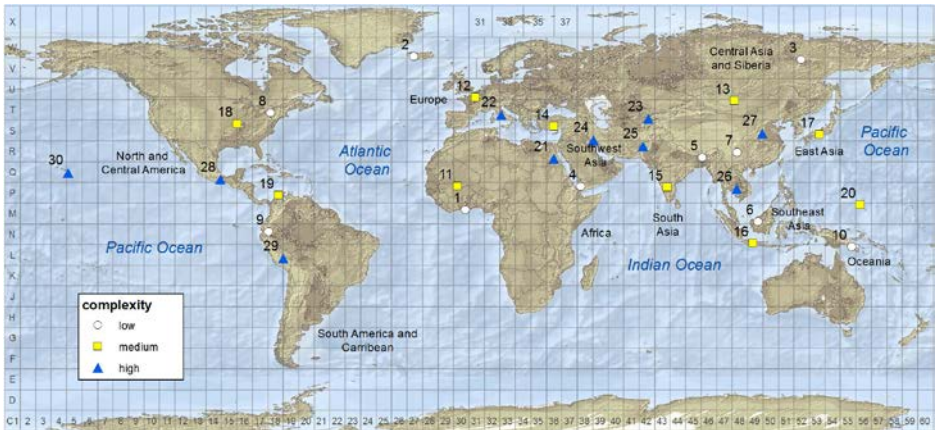

Figure 2. Locations of the 30 sampling points (Natural Geographic Areas) on the world map. For the key to NGA numbers, see Table 2.

Currently, we have two classes of Territory: Natural Geographic Region and Free Form Area.

Natural Geographic Region (NGA). This type of unit is primarily a unit of data collection or analysis and is defined spatially by the area enclosed within a boundary drawn on the world map. It does not change with time. Its rough spatial scale is 100 km × 100 km (but can vary several-fold). Examples: Latium, Upper Egypt, Middle Yellow River Valley. It is the basic geographical sampling unit.

Free Form Area (FFA). This type of unit is defined spatially by the area enclosed within a boundary drawn on the world map. It can have any dimensions, and does not have to be contiguous. The purpose of FFAs is to tie various characteristics of societies and organizations to a specific set of geographic coordinates. For example, it is used to indicate what territory was controlled by a certain polity, e.g. Roman Empire, during a particular period of time (from start year to end year). Another

example is the area within which the practice of iron smelting was known (*exists within*) during a certain period of time.

Table 2. The World Sample-30. The numbers of NGAs correspond to the numbers in Figure 2.

World Region		Low Complexity		Medium Complexity		High Complexity
Africa	1	Ghanaian Coast	11	Niger Inland Delta	21	Upper Egypt
Europe	2	Iceland	12	Paris Basin	22	Latium
Central Eurasia	3	Lena River Valley	13	Orkhon Valley	23	Sogdiana
Southwest Asia	4	Yemeni Coastal Plain	14	Konya Plain	24	Susiana
South Asia	5	Garo Hills	15	Deccan	25	Kachi Plain
Southeast Asia	6	Kapuasi Basin	16	Central Java	26	Cambodian Basin
East Asia	7	Southern China Hills	17	Kansai	27	Middle Yellow River Valley
North America	8	Finger Lakes	18	Cahokia	28	Valley of Oaxaca
South America	9	Lowland Andes	19	North Colombia	29	Cuzco
Oceania-Australia	10	Oro, PNG	20	Chuuk Islands	30	Big Island Hawaii

Currently, we have the following classes of Organization:

Polity. Polity is defined as an independent political unit. Kinds of polities range from villages (local communities) through simple and complex chiefdoms to states and empires. A polity can be either centralized or not (e.g., organized as a confederation). What distinguishes a polity from other human groupings and organizations is that it is politically independent of any overarching authority; it possesses sovereignty. Polities are defined spatially by the Territories (NGAs and FFAs) with which they have spatial relationships, e.g., *controls*. Polities are dynamical entities and, therefore, their geographical extent may change with time. Thus, typically each polity will be related to a set of territories, each controlled for a specified period of time. For prehistoric periods and for NGAs populated by a

multitude of small-scale, not truly independent polities, we use a variant of Organization called a *quasi-polity* (see below).

Quasi-polity. The polity-based approach is not feasible for those periods when a NGA is divided up among a multitude of small-scale polities (e.g., independent villages or many small chiefdoms) or when it is controlled in quick succession by a number of different regimes. In instances such as these we use the concept of 'quasi-polity' (in either the spatial or temporal sense). The idea is to collect data for the quasi-polity as a whole. This way we can integrate over (often patchy) data from different sites and different polities to estimate what a 'generic' polity was like. Accordingly, when coding, for example, social complexity variables, enter data not for the whole NGA but for a 'typical' polity in it. For example, when coding for a quasi-polity, *polity territory* is not the area of the NGA as a whole, but the average or typical area of polities within the NGA. Similarly, for societies known only archaeologically we may not be able to establish the boundaries of polities, even approximately. Quasi-polity is defined as a cultural area with some degree of cultural (including linguistic, if known) homogeneity that is distinct from surrounding areas. For example, the Marshall Islands before German occupation had no overarching native or colonial authority (chiefs controlled various subsets of islands and atolls) and therefore it was not a polity. But it was a quasi-polity because of the significant cultural and linguistic uniformity.

Religious System (RS). This type of Organization is defined in ways that are analogous to a polity, except it reflects religious, rather than political authority. Religious systems are dynamical and are typically defined by a set of dated boundaries. Religious systems are more likely to overlap in space than Polities.

City. Cities are represented by a single point on the map that doesn't change with time. Although it is possible to reflect their spatial expansion dynamically, we chose not to do so in the current implementation.

Interest Group (IG). An IG is a social group that pursues some common interest, so that its members are united by a common goal or goals. Religious systems are also interest groups, but the IG category is broader. It also includes ethnic groups, professional associations, warrior bands, solidarity associations, mutual aid societies, firms and banks (including their pre-modern variants), etc. The IG is defined sociologically, not geographically. However, if desired, a territory may be associated with it in the same way as with a polity or a RS.

Sub-Polity. A *sub-polity* is an area within a polity for which variable values differ from the overarching polity. It is a general modeling feature for capturing situations in which, for example, provinces or regions within a polity have

significant differences in social organization from the rest. A sub-polity is basically a polity that lacks sovereignty.

As the databank evolves, and more questions are posed, it is conceivable that other kinds of Territory and Organization could be added.

Additionally, we recognize that most historical polities did not have sharp boundaries. Instead, in many cases, their spheres of control gradually declined from the political center, or centers. There also could be effectively stateless lacunae inside the territory nominally claimed by a state. The initial implementation of the databank does not reflect such 'facts on the ground,' but the information architecture supports it through the sub-polity class—an area that is related to a specific territory, which differs in significant ways from the rest of the polity.
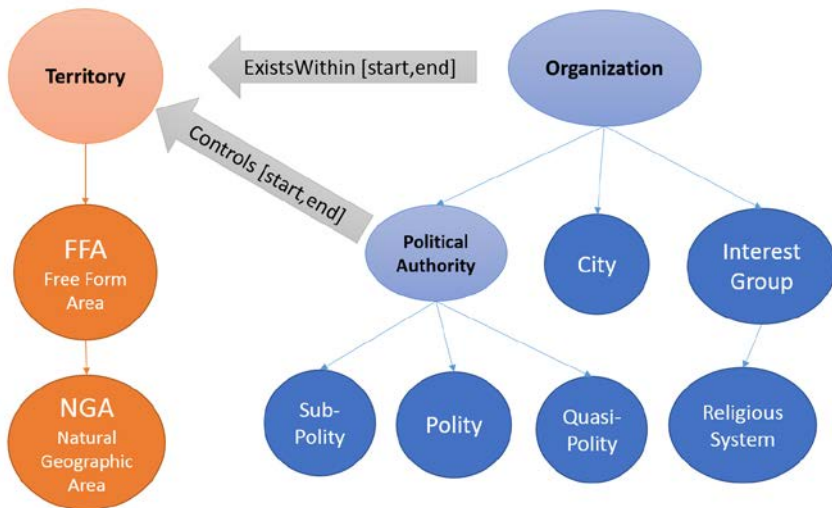

Figure 3. Seshat Meta-model, showing entity class hierarchy and the most important relationships between entity classes (note: classes with black text are abstract and thus never directly coded; they define sets of variables that are common to all sub-classes). Thin arrows indicate subclass relationships. For example, FFA is a subclass of Territory, and NGA is a subclass of FFA. Thick arrows are examples of relationships.

As a coding convention, we prefer to limit the temporal extent of polities to approximately two or three centuries, because polities tend to evolve with time and often experience transformative events; thus, it is natural to code the next

period as a different polity. For example, we divide Roman history into the following polities: Regal; Early, Middle, and Late Republic; Principate, and Dominate.

The current version of the Seshat Code Book has been developed primarily for historical societies. Coding data for societies that are known only archaeologically poses an additional set of challenges. We are currently developing an archeological Seshat Code Book that will address these challenges (Marciniak et al., forthcoming; Palmisano et al. forthcoming).

Coding Procedure

In populating the databank, the following instructions are provided to coders (using approaches based on NGAs as an example):

- Identify an NGA within the larger World Region. This should be an area ideally around 100 by 100 km, or 10,000 sq. km. Dimensions, of the NGA, however, are allowed to vary quite substantially.
- In the NGA data page, list chronologically all polities that were located in the NGA, or encompassed it (see Latium as an example). For periods when the NGA was fragmented among many small-scale polities, use the quasi-polity approach. In the intermediate case, when there were several large polities (for example, the NGA was on a frontier between two states), focus on the one that controlled the largest proportion of the NGA.
- As a coding convention, we try not to have too-long chunks of time on the same polity data sheet. Try to limit the length to 200–300 years, but at the same time don't slice it too thinly. We aim at roughly 200 (100–300) year chunks, but are guided by actual historical events that result in major change. As an example, we have split the Rome-Republican Period into three polities (Early, Middle, and Late Republic).
- Next, switch to the data page for the polity and code all polity-based variables there. In other words, you don't put any polity-related codes within the NGA sheet. The NGA is used as purely a sampling scheme, and all codes go into the relevant polity sheets.
- When coding NGA-based variables (such as resources, agriculture, and population), list chronologically the general 'epochs' and then code these variables for each period separately.
- The same approach is used to add lists of religious systems and cities, in which case entry is linked to the page for the RS or City.

Transforming Raw Data into 'Facts'

To design statistical tests of hypotheses such as those listed above, we need to define and operationalize such concepts as *social scale*, *inequality*, and *intensity of military competition*. In order to do this, we need to collect and code data

systematically and we need to be aware of the complications and complexities involved in such an endeavor. For each area of interest, we have developed a coding sheet that outlines the particular variables to be coded and how they are to be categorized or quantified. These coding sheets have been developed in conjunction with expert historians and archaeologists who have given valuable feedback on how best to classify these phenomena based on their specialist knowledge of these issues. These coding sheets are derived from our overall Codebook, which also discusses some of the complexities involved in coding data of this kind. The current version of the Seshat Codebook is included in the Supplementary Materials.

The historical and archaeological records for even the best-attested societies are incomplete, so we need to deal with the issue of missing data. Therefore, for each variable of interest, we collect data on a number of different measures, and a certain degree of redundancy in these variables is implemented by design. For many past societies (some of which are known only archaeologically), we will not be able to code every variable. Thus, different variables can serve as proxies for the same underlying factor, enabling us to compare different societies even in the face of missing data. For example, estimating populations of historical states and empires is a notoriously difficult problem; as a result, we developed a number of proxies that correlate with population numbers (e.g., the size of the largest urban center, the extent of territory controlled, etc.). Furthermore, our statistical analyses are designed to cope with missing data.

Socio-cultural phenomena described by terms such as 'social scale' or 'inequality' are often actually multidimensional and cover a range of related, but quite distinct, aspects of the human condition. Confusion or misunderstandings can sometimes occur when researchers have different definitions or conceptions of a particular term that is being used. Intriguingly, Seshat has the potential to uncover that the different dimensions of these aspects of societies may have evolved somewhat independently. For example, structural and reproductive inequality may have seen reductions since the Axial Age, but economic inequality may have been unaffected or even increased.

One strength of our approach is that we make explicit our assumptions about how we define and measure these variables. Different researchers may have different ideas about how these phenomena should be coded, or may have different interpretations about the information available in the historical and archaeological records. These different viewpoints can be readily incorporated in our scientific approach by examining them to see if these different assumptions fundamentally affect the conclusions we can draw from our investigations. For example, where different researchers propose different values for a particular variable in a particular society we can run a series of analyses to see if the main finding is robust to these alternative values.

Seshat Wiki: an Initial Implementation of the Databank

During 2012, our group developed a flexible architecture for this databank. The databank has two distinct aspects: textual/descriptive sections (including references), and coded data. The text-based sections describe what is known about a particular variable based on previous scholarship. These 'thick' descriptions provide important context about the variables being addressed, the sources of information used, and make explicit how a decision was made on a particular coding (for a more extended discussion of these issues, see Hoyer and Manning, forthcoming). The coded data are based on the narrative sections, and enable this information to be translated into a form that is amenable to statistical analysis. The databank therefore combines the best features of traditional humanistic and scientific approaches to investigating the past. The databank is implemented as a 'Wiki' (i.e., a web-based application that enables an open access, collaborative project), which allows text on different variables to be easily added and updated. The information from the Wiki is 'scraped' automatically and translated into formats (e.g., tab-delimited files) that can be analyzed in statistical software packages.

Although the Wiki implementation was a suitable approach during the initial phases of the databank's development, it is rapidly becoming a limiting factor due to the growth of data that is added daily. As of the writing of this paper, we have developed a plan to port the Wiki data into an RDF-based triplestore (Schreiber and Raimond 2014). This type of graph-based representation is particularly suitable for capturing rich, structured knowledge about complex entities and multi-faceted relationships between them. In contrast with a traditional SQL-based relational database, graph-based knowledge models, specified in OWL (the web ontology language) or RDFS (resource description framework schema), facilitate constant evolution of schema. In addition to the underlying triplestore, we will use cutting-edge techniques developed by Kevin Feeney, Rob Brennan, and colleagues to further facilitate data collection, curation and use, as described in the next section.

Seshat Databank: Overarching Plan

The design goal is to maximize the research community's collective intelligence while minimizing overheads. Ultimately, we aim to create a system with easy-to-use software tools that support the following roles and features:

Seshat Contributors. Non-technical users, such as historians and archaeologists, can easily add data to the system and update existing data (e.g., using graphical

input tools). Ultimately, we would like to make this open and robust enough so that any web user can suggest updates to the datasets in an easy way (crowd-sourcing).

Seshat Editors. Data administrators can moderate, correct, and manage the data in the system over time. The downside to harvesting data from a broad community is that inaccuracies, disagreements, user errors, and malicious use must be managed or else the dataset will degrade in quality over time.

Seshat Data Architects. Knowledge engineers can make changes to the schema (data structures) over time and manage transitions between versions of the schema without breaking databank integrity (maintaining and assuring the accuracy and consistency of data over its entire life-cycle).

Seshat Analysts. Statisticians and mathematical modelers will prepare and analyze time series data to investigate big questions about human societies.

Seshat Readers. General end-users can browse, search, download, and view arbitrary slices of the data in a very wide range of attractive and helpful ways.

Seshat Administrators. Technical administrators can manage the data curation and publication platform or servers to deal with changes in data, schemata, collection tools and publication formats or tools (e.g., visualizations) over time in a scalable fashion.

Seshat will be managed through the Dacura Linked Data platform (http://dacura.scss.tcd.ie), developed at Trinity College Dublin. Dacura provides support for dataset capturing, curation, and publication (Feeney 2014). Nonetheless, when dealing with a complex and multi-faceted domain, the development of formal schema and tools to facilitate convenient and accurate data input requires considerable experimental and developmental effort. The nature of the Seshat data is such that there are many opportunities to take advantage of maps and timelines to capture the spatial and temporal aspects of the data. However, these tasks are labor-intensive, so the development of the databank system will proceed incrementally. The Seshat databank will be progressively migrated from the current Wiki to the Dacura Linked Data platform in a number of phases, each designed to add functionality to the system, progressively increasing its ability to gather and present high quality structured data, without interrupting researchers' ability to add new data. The following text describes the initial phases in the migration.

Phase 1: Data Validation and Extraction, Seshat Data Web

The first phase of development will retain the current Wiki as the authoritative source of data and the data-input tool. It will develop a RDF schema (or knowledge model), which will be used to validate the data in the Wiki and prepare the ground for future migration. This involves two parallel development tasks.

First, we must develop the first version of an RDF-based Seshat Knowledge Model, including definition of vocabularies and formal specification of relationships between concepts. This will be an incremental process, starting with the core concepts and working outwards to the details.

Second, we must develop two tools which (a) extract the structured, natural language data from the wiki into a time series data dump amenable to statistical data analysis and (b) formally validate the data in the wiki to eliminate coding errors and remove ambiguity. These two tools enable us to perform analysis on the data; to track the size and growth of the databank; to detect inconsistencies in the data set; to copy the data from the wiki to a complementary graph-based database (RDF triplestore); and to generate data problem reports for the Seshat Contributors and Editors. The data problem reports will enable Editors or Contributors to correct and improve the data in the wiki and the Seshat Code Book.

In addition to these tasks, we will use the existing facilities of the Dacura platform to perform more extensive (semantic) data quality checks, using the RDF schema to ensure that any updated data conforms to the knowledge model, and publish the RDF dataset as a simple Linked Data site. This system will operate in parallel to the existing Wiki so current work practices and data ownership can continue. Figure 4 illustrates the publication pipeline of the Seshat system, as it exists in the first phase.

Progress. We have completed most of the development of this phase. Including (1) a first draft of the Seshat RDF Knowledge Model (Figure 2 shows the high-level meta-model) and (F2) a working software system, which extracts the variable data from the Seshat Wiki and analyzes it against the Code Book, producing reports where there are non-compliant data entries and generating TSV-formatted extractions of the data for analysis.

These outputs are currently operational in the Seshat data collection system. This allowed us to create a Wiki-based Data Validation tool that is integrated into each wiki page, allowing coders (Seshat Contributors) to check their work for consistency as they enter data. In addition, our Data Export tool allows Seshat Editors to view the data as it is being compiled and Seshat Analysts to begin to work on modeling problems.

Further Phases. The goal is to progressively replace the Wiki input format with graphical user interface tools which use the schema definition to maintain data quality over time and to ease the task of creating and updating complex instances of data types. We will progressively add enhancements to the publishing capabilities—developing richer interfaces for browsing, visualizing, analyzing, and interlinking the Seshat databank with external sources of knowledge. Once this

process is sufficiently advanced, we will switch the authoritative source of data from the wiki to the triplestore, retaining the Wiki as a publishing outlet.

The Future of Seshat

As described in the previous section, at the time of writing this paper, the main focus of the Seshat Databank development team is on coding the NGAs in the World Sample-30 for a set of variables defined by the current list of Big Questions. The variable classes include social complexity, warfare, rituals, agriculture and resources, institutions and equity, and economics and well-being (see the current version of the Code Book in the Supplementary Online Materials for a detailed description of the variables). Simultaneously, we are making a transition from the current Wiki to the Seshat triplestore managed through the Dacura Linked Data platform. Accomplishing these near-term goals will provide us with a solid basis from which to progressively expand the Databank in multiple mutually supportive directions.
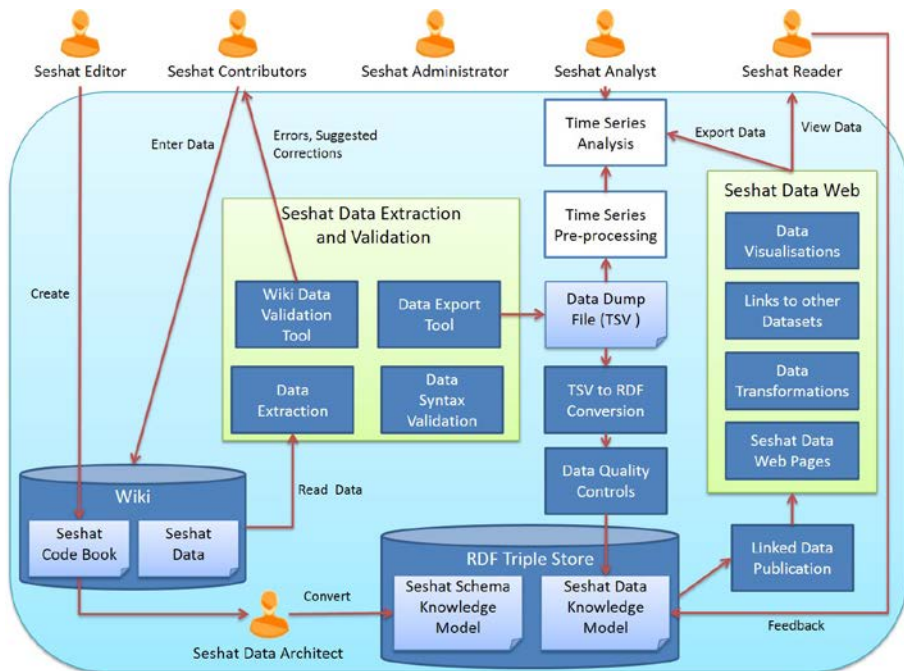

Figure 4. Seshat Phase 1 publishing pipeline architecture.

The two most important dimensions of Seshat expansion in the medium term are (1) adding more variables and (2) increasing the databank coverage to a

progressively greater fraction of the world's surface. As far as new variables are concerned, it would be extremely interesting to add data on the evolution of technology and on linguistic evolution. These variable classes are interesting in themselves because each is associated with a developed body of theory (for recent reviews, see Boyd et al. 2013, Gray et al. 2013), and they can be important *explanans* variables in helping us explain other aspects of the cultural evolution of human societies. Another exciting area of research is gene-culture coevolution, which has, so far, been primarily investigated with theoretical models. There is little reason to doubt that the ongoing methodological advances in ancient DNA will soon enable us to test these theories with data (Callaway 2015).

Increasing the thematic and geographic coverage, however, will come at a cost. Projecting from the current rate of data accumulation, we estimate that by 2017, the size of the Seshat databank will approach—or exceed—the symbolic threshold of one million 'facts.'[3] But even this enormous amount of data will be restricted in its thematic and, especially, geographic coverage. It is already clear that extending the geographic coverage to the whole world is not feasible using our current data collection approach.

One direction that we are currently exploring is crowd sourcing—developing software that will support recruitment of volunteers to assist in manual data processing. More radically, we will need to transition to a technology that automates the harvesting of the required variable values from open web-sources, including, but not limited to, the repositories of scholarly publications. Thus, we intend to explore the possibility of harvesting data from such source as JSTOR's archive of academic publications.

While harvesting data from unstructured massive repositories, such as JSTOR, is a future direction of Seshat development, there are also numerous sources of structured data available, which contain data relevant to the Seshat Databank. The Linked Data technology, on which the Dacura platform is based, facilitates interlinking the Seshat data with other datasets and supports applications and analysis that integrates Seshat data with other sources, as long as they also use the RDF technology. Given the richness of the Seshat databank and the expanding web of data, there will be many important insights to be gained by interlinking data that go beyond the scope of the Seshat big questions themselves.

Another very important direction in which to develop is improved data quality. Once the databank has been migrated to the triplestore, each data point in it will have searchable, structured provenance records associated with it, which will detail the precise source of each variable value. This will go beyond logging information on data changes and will connect data values with the source of

---

[3] We are using 'facts' here to refer to individual RDF triples. A typical Seshat 'data point,' depending on its informational complexity (such as temporal extent, presence of uncertainty or disagreement, etc.), is encoded with 3–10 triples.

academic research or primary materials. For example, we can have multiple alternative estimates of the population of some polity at a certain point in time: one from a contemporary source, such as a tax survey; others based on opinions of different experts, which may strongly disagree with one another; and yet another based on the estimate of the maximum carrying capacity given the contemporary agricultural technology. With this information, we will be able to link values with evidence, supporting informed debate or a process for resolving disagreements or ambiguities.

Finally, the value of the Databank will be enhanced by developing support for applying different interpretative frameworks to the underlying data, supporting multiple different views of history, perhaps associated with a particular expert, research team, or research methodology. For example, an estimate of the carrying capacity, which feeds into the estimated population, itself depends on many assumptions of the models used to calculate it. The impact of using alternative assumptions in the models or theories, leading to different interpretations, can then be quickly analyzed and evaluated. The next logical step is to implement a catalog of simulation models, which each can generate predictions about the data based on different mechanisms. As the background data become more accurate and plentiful, simulations could then be re-run by any Seshat user to discover how such improved or alternative data affect model predictions. This novel capability will lead to a more rapid process of theory evaluation, and to quantification of the limits for the application of a given theory.

Despite the current excitement associated with the digital humanities, historians and social scientists have not really begun to utilize the full power of what modern Information Technology can deliver. We believe that the new IT capabilities will eventually (and in the not-too-distant future) transform the field of historical social sciences into what the sociologist Randall Collins termed *rapid-discovery science* (Collins 1994, although it should be noted that Collins himself was sceptical that such transformation is likely). It is our hope that Seshat: the Global History Databank will be one of the key mechanisms by which such a transformation will be effected.

Acknowledgments

contributions of our team of research assistants, post-doctoral researchers, consultants, and experts. Additionally, we have received invaluable assistance from our collaborators. Please see the Seshat website for a comprehensive list of private donors, partners, experts, and consultants and their respective areas of expertise.

References

Algaze, G. 2008. Ancient Mesopotamia at the Dawn of Civilization: The Evolution of an Urban Landscape. University of Chicago Press, Chicago.

Ames, K. 2007. The Archaeology of Rank. Pages 487-513 *in* R. A. Bentley, H. D. G. Maschner, and C. Chippendale, editors. Handbook of Archaeological Theories. Alta Mira Press, Lanham.

Atkinson, Q. D., and H. Whitehouse. 2010. The cultural morphospace of ritual form: Examining modes of religiosity cross-culturally.

Axelrod, R., and W. D. Hamilton. 1981. The evolution of cooperation. Science 211:1390-1396.

Bar-Yosef, O. 2001. From Sedentary Foragers to Village Hierarchies: the Emergence of Social Institutions. Pages 1-38 *in* G. Runciman, editor. The origin of Human Social Institutions. British Academy, London.

Bellah, R. N. 2011. Religion in Human Evolution: From the Paleolithic to the Axial Age. Harvard University Press, Cambridge, MA.

Boehm, C. 2001. Hierarchy in the Forest: The Evolution of Egalitarian Behavior. Harvard University Press, Harvard.

Boehm, C. 2012. Moral Origins: The Evolution of Virtue, Altruism, and Shame. Basic Books, New York.

Borgerhoff Mulder, M., S. Bowles, T. Hertz, A. Bell, J. Beise, G. Clark, I. Fazzio, M. Gurven, K. Hill, P. L. Hooper, W. Irons, H. Kaplan, D. Leonetti, B. Low, F. Marlowe, R. McElreath, S. Naidu, D. Nolin, P. Piraino, R. Quinlan, E. Schniter, R. Sear, M. Shenk, E. A. Smith, C. v. Rueden, and P. Wiessner. 2009. Intergenerational Wealth Transmission and the Dynamics of Inequality in Small-Scale Societies. Science 326:682-688.

Bowles, S. 2009. Did Warfare Among Ancestral Hunter-Gatherers Affect the Evolution of Human Social Behaviors? Science 324:1293-1298.

Boyd, R., P. J. Richerson, and J. Henrich. 2013. The Cultural Evolution of Technology: Facts and Theories. Pages 119-142 *in* P. J. Richerson and M. H. Christiansen, editors. Cultural Evolution: Society, Technology, Language, and Religion. MIT Press, Cambridge, MA.

Breasted, J. H. 1919. The Oriental Institute of the University of Chicago. American Journal of Semitic Languages and Literatures 35:196-204.

Callaway, E. 2015. DNA deluge reveals Bronze Age secrets: Population-scale studies of ancient genomes hint at roots of technology, languages and diet. Nature 522:140-141.

Carballo, D. M., P. Roscoe, and G. M. Feinman. 2014. Cooperation and Collective Action in the Cultural Evolution of Complex Societies. J Archaeol Method Theory 21:98-133.

Carneiro, R. L. 1970. A theory of the origin of the state. Science 169:733-738.

Carneiro, R. L. 1998. What happened at the flashpoit? Conjectures on chiefdom formation at the very moment of conception. Pages 18-42 *in* E. M. Redmond, editor. Chiefdoms and chieftaincy in the Americas. University of Florida Press, Gainsville.

Cashdan, E. A. 1980. Egalitarianism among Hunters and Gatherers. American Anthropologist 82:116-120.

Collins, R. 1994. Why the social sciences won't become high-consensus, rapid-discovery science. Sociological Forum 9:155-177.

Currie, T. E., A. Bogaard, R. Cesaretti, N. Edwards, P. Francois, P. Holden, D. Hoyer, A. Korotayev, J. Manning, J. C. M. Garcia, O. K. Oyebamiji, C. Petrie, P. Turchin, H. Whitehouse, and A. Williams. 2015. Agricultural productivity in past societies: Toward an empirically informed model for testing cultural evolutionary hypotheses. Cliodynamics.

Dubreuil, B. 2008. Strong Reciprocity and the Emergence of Large-Scale Societies. Philosophy of the Social Sciences 38:167-191.

Dunbar, R. I. M. 1992. Neocortex size as a constraint on group size in primates. Journal of Human Evolution 22:469-493.

Dunbar, R. I. M., and S. Shultz. 2007. Evolution in the social brain. Science 317:1344-1347.

Fehr, E., and S. Gächter. 2002. Altruistic punishment in humans. Nature 415:137-140.

Feinman, G. M., and J. Marcus. 1998. Archaic States. School of American Research Press, Santa Fe.

Flannery, K., and J. Marcus. 2012. The Creation of Inequality: How Our Prehistoric Ancestors Set the Stage for Monarchy, Slavery, and Empire. Harvard University Press, Cambridge, MA.

Fried, M. H. 1967. The evolution of political society: An essay in political anthropology. Random House, New York.

Gray, R. D., S. J. Greenhill, and Q. D. Atkinson. 2013. Phylogenetic Models of Language Change: Three New Questions. Pages 285-300 *in* P. J. Richerson and M. H. Christiansen, editors. Cultural Evolution: Society, Technology, Language, and Religion. MIT Press, Cambridge, MA.

Hayden, B. 1995. Pathways to power: principles for creating socioeconomic inequalities. Pages 15-86 *in* T. D. Price and G. Feinman, editors. Foundations of social inequality Plenum, New York.

Hempel, C. G., and P. Oppenheim. 1948. Studies in the Logic of Explanation. Philosophy of Science 15:135-175.

Henrich, J., J. Ensminger, R. McElreath, A. Barr, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker. 2010. Markets, Religion, Community Size, and the Evolution of Fairness and Punishment. Science 327:1480-1484.

Ibn Khaldun. 1958. The Muqaddimah: An Introduction to History. Translated from the Arabic by Franz Rosenthal. Pantheon Books, New York.

Jaspers, K. 1953. The Origin and Goal of History. Routledge & Kegan Paul, New York.

Johnson, A. W., and T. Earle. 2000. The evolution of human societies: from foraging group to agrarian state, 2nd edition. Stanford University Press, Stanford, CA.

Keely, L. H. 1997. War Before Civilization: The Myth of the Peaceful Savage. Oxford University Press, New York.

Kirch, P. V. 2010. How Chiefs Became Kings: Divine Kingship and the Rise of Archaic States in Ancient Hawai'i. University of California Press, Berkeley.

LeBlanc, S. A. 2003. Constant Battles: Why We Fight. St. Martin's Griffin, New York.

Lee, R. B., and I. DeVore. 1968. Problems in the study of hunters and gatherers. Pages 3-12 *in* R. B. Lee and I. DeVore, editors. Man the Hunter. Aldine, Chicago.

Liverani, M. 2006. Uruk: The First City. London and Oakville:. Equinox, London.

Mann, M. 1986. The sources of social power. I. A history of power from the beginning to A.D. 1760. Cambridge University Press, Cambridge, UK.

Maynard Smith, J., and E. Szathmáry. 1995. The Major Transitions in Evolution. W. H. Freeman, New York.

Morrison, K. D. 1994. The Intensification of Production: Archaeological Approaches. Journal of Archaeological Method and Theory 1:111-159.

Mullins, D., H. Whitehouse, and Q. Atkinson. 2013. . Journal of Economic Behavior and Organization 90:S141-S151.

Norenzayan, A., A. F. Shariff, A. K. Willard, E. Slingerland, W. M. Gervais, R. A. Mcnamara, and J. Henrich. 2015. The Cultural Evolution of Prosocial Religions. Behavioral & Brain Sciences.

Oppenheimer, F. 1975. The State; Its History and Development Viewed Sociologically. Free Life Editions, New York.

Price, T. D., and G. Feinman, editors. 1995. Foundations of Social Inequality. Springer, New York.

Richerson, P. J., and R. Boyd. 1998. The evolution of human ultrasociality. Pages 71-95 *in* I. Eibl-Eibesfeldt and F. K. Salter, editors. Ethnic Conflict and Indoctrination. Berghahn Books, Oxford.

Richerson, P. J., and R. Boyd. 2001. The evolution of subjective commitment to groups: a tribal instincts hypothesis. Pages 186-220 *in* R. M. Nesse, editor. Evolution and the capacity for commitment. Russel Sage Foundation, New York.

Sanderson, S. K. 1999. Social Transformations: a General Theory for Historical Development. Rowman and Littlefield, Lanham, MD.

Scheidel, W. 2013. Studying the State. Pages 5-57 *in* P. F. Bang and W. Scheidel, editors. The Oxford Handbook of the State in the Ancient Near East and Mediterranean. Oxford University Press, Oxford.

Schreiber, G. and Y. Raimond. 2014. RDF 1.1 Primer. W3C Working Group Note 25 February 2014.

Service, E. R. 1975. Origins of the State and Civilization: The Process of Cultural Evolution. Norton, New York.

Trigger, B. G. 2003. Understanding Early Civilizations. Cambridge University Press, Cambridge.

Turchin, P. 2009. A Theory for Formation of Large Empires. Journal of Global History 4:191-207.

Turchin, P. 2011. Warfare and the Evolution of Social Complexity: a Multilevel Selection Approach. Structure and Dynamics 4(3), Article 2:1-37.

Turchin, P. 2013. The Puzzle of Human Ultrasociality: How Did Large-Scale Complex Societies Evolve? Pages 61-73 *in* P. J. Richerson and M. H. Christiansen, editors. Cultural Evolution: Society, Technology, Language, and Religion. MIT Press, Cambridge, MA.

Turchin, P., T. E. Currie, E. A. L. Turner, and S. Gavrilets. 2013. War, Space, and the Evolution of Old World Complex Societies. PNAS, published online Sept. 23, 2013. PDF. PNAS 110:16384–16389.

Turchin, P., and S. Gavrilets. 2009. Evolution of Complex Hierarchical Societies. Social History and Evolution 8(2):167-198.

Whitehouse, H. 2004. Modes of religiosity: a cognitive theory of religious transmission. AltaMira Press, Walnut Creek.

Whitehouse, H., and I. Hodder. 2010. Modes of Religiosity at Çatalhöyük.*in* I. Hodder, editor. Religion in the Emergence of Civilization: Çatalhöyük as a case study. Cambridge University Press, Cambridge.

Whitehouse, H., and J. A. Lanman. 2014. The Ties That Bind Us: Ritual, Fusion, and Identification. Current Anthropology 55:674-695.

Wilkinson, T., G. Philip, J. Bradbury, R. Dunford, D. Donoghue, N. Galiatsatos, D. Lawrence, A. Ricci, and S. Smith. 2014. Contextualizing Early Urbanization: Settlement Cores, Early States and Agro-Pastoral Strategies in the Fertile Crescent during the Fourth and Third Millennia BC. Journal of World Prehistory 27:43-109.

Wilson, D. S. 2002. Darwin's Cathedral: Evolution, Religion, and The Nature of Society. University of Chicago Press, Chicago.

Zeder, M. A. 2011. The origins of agriculture in the Near East. Current Anthropology 52:221-235.