

BIG DATA I SIECI SEMANT

BAZA SESHAT W STUDIACH NAD

Ostatnie lata przynoszą ogromny przyrost danych naukowych, które są coraz łatwiej dostępne w sieci. Dzieje się to głównie za sprawą rozwoju nowoczesnych baz danych, programów digitalizacji archiwów i zbiorów muzealnych czy wreszcie polityki instytucji finansujących badania naukowe i inicjatywy kulturalne, wymagających aby różnego rodzaju materiały, analizy i wyniki badań były upowszechniane w wolnym dostępie.

**PATRYCJA FILIPOWICZ,
ARKADIUSZ MARCINIAK**



Seshat

Logo projektu Seshat



FIGURKA NEOLITYCZNEJ „BOGINI PŁODNOŚCI” ODKRYTA PRZEZ POZNAŃSKICH ARCHEOLOGÓW W ÇATALHÖYÜK (TURCJA). PODOBNE FIGURKI ZNJDOWANE SĄ NA OSADACH PIERWSZYCH ROLNIKÓW W EUROPIE POŁUDNIOWEJ I ŚRODKOWOSCHODNIEJ.

YCZNE W ARCHEOLOGII ZRÓŻNICOWANIEM CZŁOWIEKA

OD BAZ DANYCH DO SIECI SEMANTYCZNYCH

Współczesność stwarza nowe wyzwania wobec historyków i archeologów, które wiążą się z koniecznością porządkowania informacji archiwalnych w procesach przekształcania danych naukowych do postaci elektronicznej i wykorzystania całości we własnym warsztacie naukowym. Te zmiany spoza obszaru nauk historycznych otworzyły zatem nieistniejące wcześniej możliwości analizy procesów dziejowych przy wykorzystaniu ogromnej ilości danych. W rezultacie powstały warunki fundamentalnych zmian w zakresie metodologii nauk historycznych związane z możliwością testowania komplementarnych wobec siebie hipotez.

WIELKIE ZASOBY DANYCH W ARCHEOLOGII

Dynamiczny przyrost danych jest szczególnie czytelny w archeologii. Istnienie wielkich, nieuporządkowanych i wciąż przyrastających cyfrowych zasobów danych (szara literatura, bazy danych, artykuły i publikacje dostępne *online*, strony internetowe, podcasty itd.) to palący problem współczesnych archeologów. Oprócz tego, że danych generowanych w toku badań jest po prostu za dużo do ich objęcia, pojawia się też kwestia ich jakości: czy są one wartościowe. Już ponad 10 lat temu zwrócono uwagę na potrzebę budowy cyfrowej infrastruktury dla archeologii (Kintigh 2006) i choć nie brakuje tego typu inicjatyw, to wydaje się, że to właśnie jakość danych powinna być kluczową kwestią do rozpatrzenia.

Pomocne w metodycznym uporządkowaniu i zarządzaniu zasobami danych jest tworzenie tzw. ontologii. O ile w filozofii ontologia to nauka o bycie, to w informatyce związanej z inżynierią wiedzy, to hierarchiczny system kategorii i relacji. Można ją zdefiniować, jest formalną specyfikacją wydzielonej konceptualizacji. Oznacza ona abstrakcyjny model zjawisk z wybranej domeny identyfikujący istotne dla niej pojęcia. Innymi słowy ontologia oznacza terminy i reguły przetwarzania wykorzystywane dla opisu wybranej dziedziny wiedzy. Formalny charakter tej konceptualizacji jest niezbędny, aby mógł być zrozumiały przez maszyny.

Dziś ontologii i rozmaitych repozytoriów informacji nie brakuje: można wśród nich wymienić np. *HRAF – Human Relations Area Files* (<http://hraf.yale.edu/>), *Open Context* (<http://opencontext.org/>), *The Digital Archaeological Record (tDAR)* (<https://www.tdar.org/>), *ARIADNE* (<http://www.ariadne-infrastructure.eu/About>) czy portal *E-Archaeology.org* (<http://e-archaeology.org/projects/project-2010-2012/objectives.html>). Ten ostatni został stworzony przez zespół z Uniwersytetu im. Adama Mickiewicza (UAM) w Poznaniu w ramach projektu europejskiego: *The Lifelong Learning programme Leonardo da Vinci. Transfer of innovation*. Powstał jako efekt tworzenia treści dydaktycznych dotyczących dziedzictwa dla różnych grup odbiorców. Jednym z ostatnich, ambitnych przedsięwzięć w zakresie budowania ontologii dziedzinowych dla archeologii, w celu sformalizowanej konceptualizacji zjawisk i połączonych z nimi pojęć, w odniesieniu do pradziejowej przeszłości, jest projekt *Seshat*, w który jest zaangażowany również UAM w Poznaniu.

CZYM JEST SESHAT?

Projekt, którego patronką została egipska bogini wiedzy i pisma – Seshat, wystartował w 2011 roku. Jego twórcami i koordynatorami są badacze związani z Uniwersytetem w Oxfordzie: Peter Turchin (propagator kliodynamiki: historii jako nauki ścisłej, w której istotne jest modelowanie matematyczne) oraz Harvey Whitehouse (przedstawiciel religioznawstwa kognitywnego i twórca teorii stylów religijności). W 2014 roku w projekt włączył się zespół z Instytutu Archeologii w Poznaniu, koordynowany przez prof. Arkadiusza Marciniaka, który został odpowiedzialnym za jego część archeologiczną.

Seshat: The Global History Databank (<http://seshatdatabank.info/>) jest innowacyjną bazą danych historycznych i archeologicznych, która ma zgromadzić w jednym miejscu najważniejsze, wieloaspektowe informacje o społecznościach z całego świata, począwszy od neolitu aż do XIX wieku. W zamyśle twórców, tak skonstruowana baza pozwoli spojrzeć na rozwój cywilizacji w ogromnej perspektywie; umożliwi badanie procesów długiego trwania oraz być może pozwoli odpowiedzieć na „Wielkie Pytania” (tzn.

O autorach:



**Prof. ARKADIUSZ
MARCINIAK**

Profesor archeologii na Uniwersytecie im. Adama Mickiewicza w Poznaniu. Jest kierownikiem badań wykopaliskowych w Çatalhöyük. Specjalizuje się w problematyce neolitu bliskowschodniego i europejskiego, zajmuje się konceptualizacją badań nad przeszłością społeczną oraz jest twórcą społecznej zooarcheologii.



PATRYCJA FILIPOWICZ

Archeolog, obecnie pracuje w Instytucie Archeologii UAM w Poznaniu w ramach europejskiego projektu ALIGNED: Quality-centric, Software and Data Engineering. Od 2005 roku uczestniczy w badaniach wykopaliskowych neolitycznego stanowiska Çatalhöyük w Turcji.



Mapa przedstawiająca regiony geograficzne kodowane w Seshacie przez zespół Instytutu Archeologii UAM.

pytania np. o pojawienie się i rozwój złożonych społeczeństw, na które odpowiedzi nie przyniosą pojedyncze wykopaliska czy projekty badawcze).

U sedna projektu *Seshat* tkwi właśnie pytanie o ewolucję złożonych organizmów społecznych, a architektura bazy danych jest odzwierciedleniem takiego podejścia ewolucyjnego. Istotą bazy jest rozpoznanie uwarunkowań, w jakich pojawiają się i rozprzestrzeniają różne formy organizacyjne i instytucje społeczne. *Seshat* ma pozwalać na wychwycenie zależności pomiędzy wzrostem złożoności a zmianami na różnych płaszczyznach (np.: w domenie gospodarczej, religijnej itd.) oraz na testowanie rozmaitych hipotez dotyczących przyczynowości zmian społeczno-kulturowych.

Według Petera Turchina najpotężniejszym czynnikiem, wyjaśniającym powstanie i ewolucję złożonych organizmów społecznych, jest współzawodnictwo, przyjmujące formę konfliktów zbrojnych. Swą hipotezę oparł na założeniu, że wraz ze wzrostem intensywności konfliktu zbrojnego, wzrasta selekcja na rzecz powstawania złożonych instytucji społecznych (grup społecznych). Intensywność konfliktu zbrojnego jest natomiast spowodowana dostępnością technologii militarnej oraz sprzyjającymi warunkami geograficznymi. Zespół Turchina postanowił przetestować tę hipotezę, dla obszaru Afroeurazji w okresie od 1500 r. przed Chr. do 1500 r. po Chr., bazując na istniejących danych empirycznych. Otrzymane wyniki były w 65%

zgodne z rzeczywistym z rozprzestrzenieniem się dużych społeczności na tym obszarze, a model ten potwierdził ważną rolę instytucji w budowaniu wczesnych organizmów państwowych (Turchin et al. 2013).

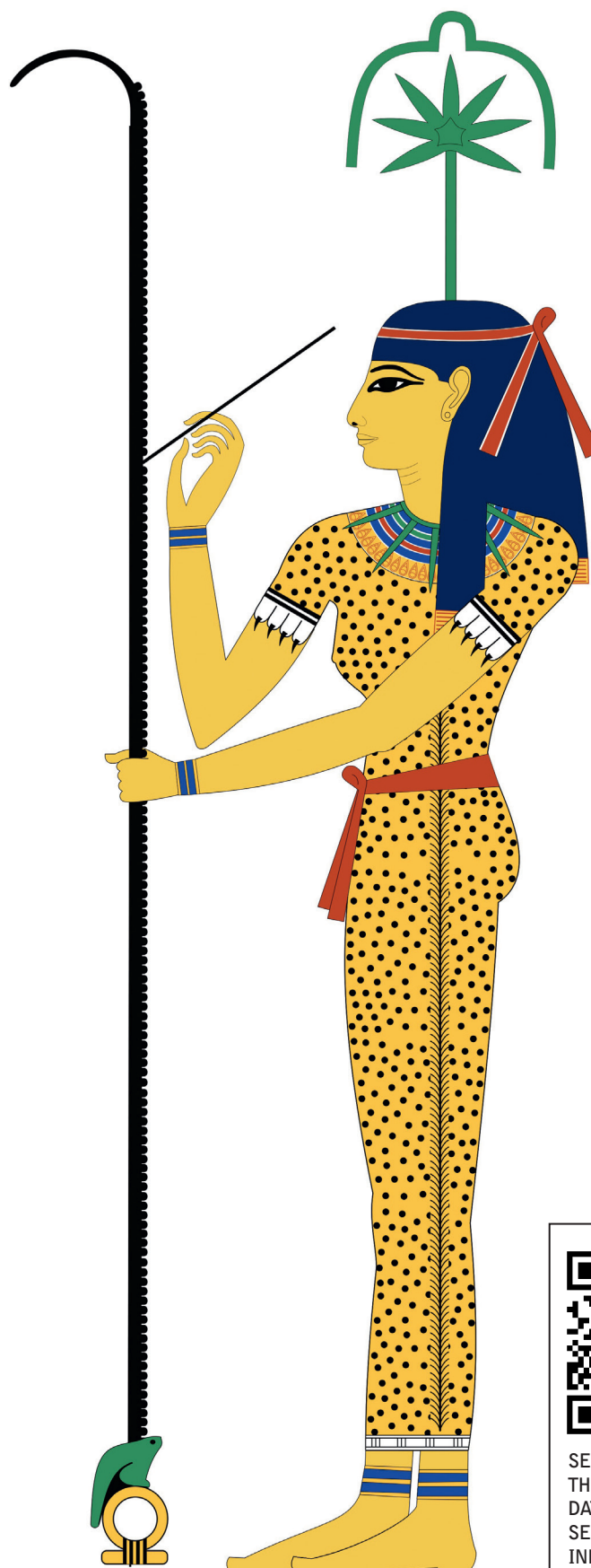
SESHAT OD KUCHNI

Architektura bazy danych jest prosta i przejrzysta: składa się z jednostek analitycznych i przyporządkowanych im szeregu zmiennych, dotyczących wielu aspektów takich jak: aspekty społeczne i kulturowe, rytuały, wojna, handel i wymiana itp. Podstawową jednostką analityczną jest tzw. NGA (Naturalny Region Geograficzny) – jest to ograniczony obszar, spójny i charakterystyczny geograficznie (np. Górny Egipt). W obrębie danego NGA wydziela się kolejne niezależne jednostki polityczne lub terytorialno-kulturowe – tzw. *polities* (ang. ustroje), a dla społeczeństw prehistorycznych: *quasi-polities* (np. Okres Starożytności w Egipcie), które formowały się w danym regionie. *Polities* w danym obszarze są uszeregowane w porządku chronologicznym. Dla każdej *polity* należy zebrać informacje dotyczące szeregu zmiennych geograficznych, osadniczych, społecznych, gospodarczych i wierzeniowych. Sposób ich zapisu został sformalizowany w postaci tzw. *codebooka*, który określa zmienne i ich atrybuty. Od 2011 roku nieprzerwanie pracuje nad tym ogromny, międzynarodowy zespół ekspertów, którzy uzupełniają arkusze na podstawie systematycznego przeglądu publikacji naukowych oraz własnej wiedzy eksperckiej.

SESHAT W SŁUŻBIE ARCHEOLOGII

U swego zarania baza została zaprojektowana z punktu widzenia potrzeb badań historycznych i antropologii kultury. Przyjęta formuła w znaczący sposób nie odpowiadała specyfice społeczności pradziejowych. Zasadniczą przeszkodą w zbieraniu informacji na ich temat jest nieistnienie źródeł pisanych. Zabrakło zmiennych, za pomocą których można by zakodować informacje uzyskane na podstawie źródeł materialnych, pozyskanych w efekcie badań wykopaliskowych. Narodził się więc pomysł stworzenia odrębnego *Seshatu* Archeologicznego. Do jego powstania prowadziły burzliwe dyskusje na czterech spotkaniach: w Oxfordzie (2014-2015) i w Santa Fe (2015).

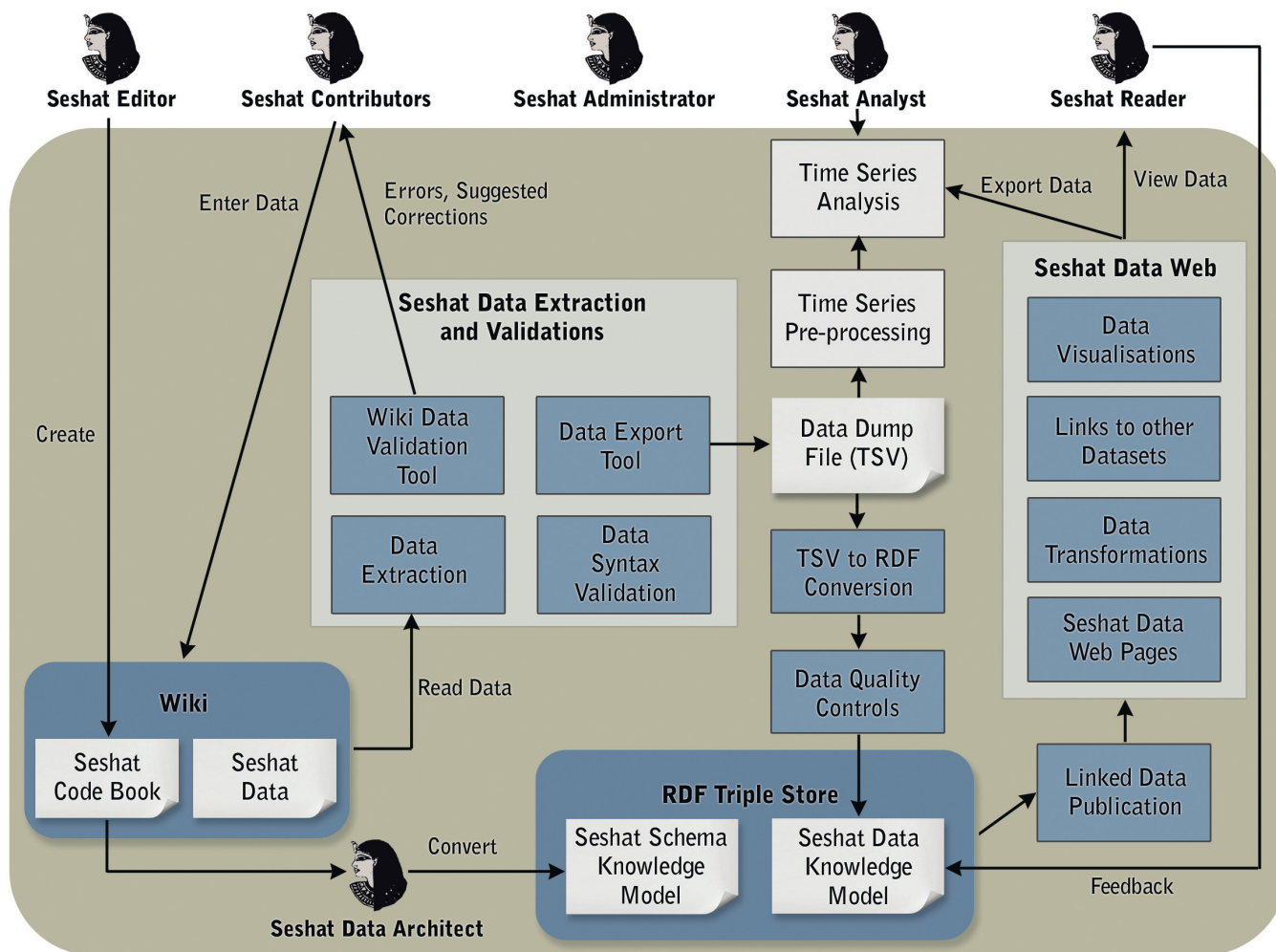
Zadaniem zespołu z poznańskiego Instytutu Archeologii było zmodyfikowanie istniejącej bazy danych w sposób umożliwiający gromadzenie danych pozyskiwanych w praktyce archeologicznej. Wymagało to stworzenia nowych kategorii i zdefiniowanie typowo archeologicznych zmiennych. Na początku 2016 roku rozpoczęliśmy pracę nad zbiorem zmiennych archeologicznych oraz arkuszami dotyczącymi poszczególnych, interesujących nas aspektów. W rezultacie, nowy



Egipska bogini wiedzy i pisma Seshat



SESHAT:
THE GLOBAL HISTORY
DATABANK (WWW.
SESHATDATABANK.
INFO)



archeologiczny *codebook* historycznej bazy *Seshat* został poszerzony od zmienne m.in. odnoszące się do nierówności społecznych (uwzględniliśmy m.in. pochówki, stanowiska archeologiczne, kontekst znalezisk itp.).

Kolejnym krokiem był wybór pierwszych NGA i *polities*. Ze względu na nasze zainteresowania badawcze i doświadczenia wykopaliskowe (zespół poznański od 2001 roku prowadzi badania neolitycznego stanowiska Çatalhöyük w Turcji), obraliśmy sobie za cel zebranie danych dotyczących społeczeństw neolitycznych na obszarze rozciągającym się od Bliskiego Wschodu do Środkowej Europy, z zamiarem prześledzenia trajektorii szeregu procesów społeczno-kulturowych i ich mechanizmów.

W tym celu wyznaczaliśmy 17 regionów geograficznych (NGA) oraz 67 jednostek *polities*, od Lewantu i Mezopotamii, poprzez Anatolię, Bałkany po Niż Polski, od wczesnego neolitu po późny chalkolit (tj. schyłek IV tys. przed Chr.). Po szczegółowym zdefiniowaniu jednostek, przystąpiliśmy do systematycznego kodowania. Na pierwszy ogień (również w celu ewaluacji

Schemat przedstawiający architekturę bazy Seshat.

zaprojektowanych przez nas arkuszy i zmiennych) wybraliśmy obszar najbliższy naszym zainteresowaniom badawczym, tzn. Anatolię. Do współpracy nad kodowaniem obszarów z kręgu bałkańskiego zaprosiliśmy badaczy neolitu z ośrodków zagranicznych (m.in. z Rumunii czy Węgier).

W najbliższym czasie zebrane przez nas dane posłużą do testowania dwóch wybranych hipotez. Pierwsza z nich dotyczy istotnych transformacji społecznych w późnym neolicie (tj. w końcu VII tysiąclecia przed Chr.), które obejmują przejście od kolektywnej, opartej na pokrewieństwie organizacji społecznej, do bardziej zindywidualizowanej formy stosunków społecznych, prowadzącej do wykształcenia się autonomicznego domostwa. Druga hipoteza dotyka kwestii powstania pierwszych *polities* w czwartym tysiącleciu przed Chr.: chcielibyśmy uzyskać odpowiedzi na szereg pytań m.in. czy *polities* powstały jako efekt intensyfikacji i formalizowania wymiany? Zebrane dane pomogą dostrzec różnice pomiędzy środkową Anatolią a północną Mezopotamią oraz odpowiedzieć



na pytania jak powstały *polities*, takie jak Halaf czy później Ubaid.

W STRONĘ SEMANTYCZNEGO SESHATU

Dynamiczny rozwój sieci semantycznych w ostatnich latach pozwolił na znaczne udoskonalenie dwóch głównych zastosowań bazy *Seshat*: porządkowania i zarządzania *big data* oraz ich przetwarzania i analizy. Pierwszym rezultatem tych zmian było przygotowanie ontologii konceptualizującej zjawiska historyczne i archeologiczne. Możliwe stało się ustrukturyzowanie i rozbudowanie formuły semantycznej, która bazując na pre-definiowanym modelu, umożliwia wykonanie automatycznej kontroli jakości wprowadzanych danych oraz ich analizy statystycznej. Sprawiało to, że dane z bazy *Seshat* zapisywane dotychczas w środowisku Wiki (na którym bazuje np. Wikipedia), zostały wyeksportowane we właściwą dla *webu* semantycznego (Internet, w którym informacje zapisane są w sposób zrozumiały dla maszyn), służącą do przedstawienia wiedzy na temat zasobów w postaci łatwo przetwarzanej przez komputer, postać *RDF (Resource Description Framework)*. Dane w takiej postaci mogły zostać poddane systematycznej ocenie pod względem ich jakości. To nowe rozwiązanie sprawia, że publikowanie danych odbywa się w zestandardyzowany sposób, a obecność automatycznie generowanego interfejsu pozwala na łatwe nimi zarządzanie. Te nowe rozwiązania informatyczne oferują także narzędzia służące rozbudowie i zarządzaniu ontologiami oraz narzędzia pozwalające na importowanie danych z innych baz danych. Potencjał tego ostatniego rozwiązania technologicznego jest trudny do przecenienia, gdyż w niedługiej perspektywie pozwoli na automatyzowane łączenie szeregu szczegółowych baz, repozytoriów czy archiwów w postać jednej zintegrowanej bazy. Wykorzystanie modelu semantycznego do publikowania i łączenia uporządkowanych danych oferuje zatem szereg niezwykle potężnych rozwiązań, przede wszystkim pozwalając na porządkowanie liczb, faktów czy stwierdzeń. Dzięki dobrze zdefiniowanej semantyce są one uporządkowane, połączone i nazwane.

Pierwszym rezultatem badawczym zastosowania rozwiązań semantycznych w odniesieniu do bazy *Seshat* była analiza, mająca na celu rozpoznanie przyczyn powstania zróżnicowania społecznego. Przedstawiający jej wyniki artykuł pt. „Quantitative historical analyses uncover a single dimension of complexity that structures global variation in human social organization” został właśnie opublikowany w *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*. Zagadnienie początków zróżnicowania społecznego jest bardzo trudne do konceptualizacji. Było to przedmiotem wielu prac i projektów badawczych, zazwyczaj opartych o bardzo małą próbę empi-

ryczną, co czyni ich wyniki problematycznymi. Korzystając z zasobów bazy *Seshat* dostępnych w formie semantycznej, dla przedmiotowej pracy wykorzystano dane z 412 *polities* pochodzących z *NGAs*. Z bazy wybrano 51 zmiennych odnoszących się do zmienności społecznej. Zmienne te zostały następnie pogrupowane w dziewięć grup opisujących kompleksowość: wielkość *polity* (I), jej zakres terytorialny (II), ilość mieszkańców „stolicy” (III), system hierarchiczny (IV), zarządzanie (V), infrastruktura (VI), system przekazywania informacji (VII), pismo (VIII) oraz system monetarny (IX). Za pomocą metody składowych głównych podjęto próbę ustalenia, czy te zmienne są ze sobą powiązane i które z nich można uznać za przyczynę zróżnicowania społecznego. Udało się ustalić, że analizowane zmienne są wyraźnie skorelowane ze sobą, co oznacza, że stopień zróżnicowania społecznego może być mierzony niezależnie przez każdą z nich. Ustalono także, że organizacja społeczna ewoluuje w przewidywalny sposób, ale istnieją zasadnicze różnice w zakresie rozpoczęcia procesu różnicowania się społecznego oraz tempa zmiany. □

Niniejszy artykuł powstał dzięki projektowi ALIGNED: Quality-centric, Software and Data Engineering (<http://aligned-project.eu/>). Projekt jest finansowany ze środków europejskiego program Horyzont 2020 (nr umowy grantowej 644055).

ŹRÓDŁA ILUSTRACJI:

str. 70: fot. Jason Quinlan; logo *Seshat* - www.seshat-databank.info

str. 72, 74: oprac. M. Dulęba, na podstawie strony www.seshatdatabank.info

str. 73: www.seshatdatabank.info

WYKAZ CYTOWANEJ LITERATURY:

KINTIGH K. 2006. *The promise and challenge of archaeological data integration*, *American Antiquity* 71(3): 567-578.

TURCHIN P., BRENNAN, R., CURRIE, T. E., FEENEY, K. C., FRANCOIS, P., HOYER, D., ... & PEREGRINE, P. 2015. *Seshat: the global history databank*, *Cliodynamics. The Journal of Quantitative History and Cultural Evolution* 6(1): 77-107

TURCHIN P., CURRIE T., TURNER E., & GAVRILETS S. 2013. *War, space, and the evolution of Old World complex societies*. “PNAS” 110(41): 16384-16389, doi:10.1073/pnas.1308825110

TURCHIN P., CURRIE T., WHITEHOUSE H., FRANCOIS P., SPENCER CH. 2017. *Quantitative historical analyses uncover a single dimension of complexity that structures global variation in human social organization*, “PNAS” 2017, opublikowano przed drukiem 21 grudnia, doi: 10.1073/pnas.1708800115

Dziś rozmaitych repozytoriów informacji nie brakuje. Obok namiary na niektóre z nich.



OPEN CONTEXT



ARIADNE



THE DIGITAL ARCHAEOLOGICAL RECORD (TDAR)



E-ARCHAEOLOGY.ORG



HUMAN RELATIONS AREA (HRAF)