

Johan Fourie

The Data Revolution in African Economic

History Improvements in computing power, increased connectivity, and more advanced analytical techniques herald the era of Big Data in fields as diverse as astronomy, economics, biology, and management. Yet, in the history profession, this data revolution has gained only limited traction. One obvious reason is the lack of Big Data from the distant past. As Eric Schmidt, the former CEO of Google, remarked in 2010, from the dawn of civilization through 2003, five exabytes of information were created. Only seven years later, that much information was being created every two days. As the saying goes, however, not all of these data are useful, and not everything that is useful is captured in these data. Indeed, historians have to distinguish what is meaningful from what is not. Too much information can be as problematical as too little.¹

Johan Fourie is Associate Professor of Economics, Stellenbosch University. He is the author of “The Remarkable Wealth of the Dutch Cape Colony: Measurements from Eighteenth-Century Probate Inventories,” *Economic History Review*, LXVI (2013), 419–448; with Erik Green, “The Missing People: Accounting for the Productivity of Indigenous Populations in Cape Colonial History,” *Journal of African History*, LVI (2015), 195–215.

The author thanks Gareth Austin, Angus Dalrymple-Smith, Pim de Zwart, Felix Meier zu Selhausen, Ewout Frankema, Leigh Gardner, Erik Green, Kris Inwood, Martine Mariotti, Alex Moradi, Nonso Obikili, Auke Rijpma, Jan Luiten van Zanden, the members of the LEAP (Laboratory for the Economics of Africa’s Past) team at Stellenbosch, and two anonymous referees for valuable input and fruitful discussion during the writing process. He received financial support from the South African National Research Foundation, the Elite Project Fund of the Faculty of Economic and Management Sciences at Stellenbosch University, the Oppenheimer Memorial Trust Sabbatical Study Fund, the HB & MJ Thom Travel Fund, and Economic Research Southern Africa, which issued an earlier version of this paper as ERSA Working Paper #555. Jan Luiten van Zanden was a most generous sabbatical-study host in Utrecht.

© 2016 by the Massachusetts Institute of Technology and The Journal of Interdisciplinary History, Inc., doi:10.1162/JINH_a_00977

1 Eric D. Feigelson and G. Jogesh Babu, “Big Data in Astronomy,” *Significance*, IX (2012), 22–25; Liran Einav and Jonathan Levin, “The Data Revolution and Economic Analysis,” NBER Working Paper No. 19035 (2013); Vivien Marx, “Biology: The Big Challenges of Big Data,” *Nature*, CDXCVIII (2013), 255–260; A.D. Howe et al., “Big Data: The Future of Biocuration,” *ibid.*, CDLV (2008), 47–50; Andrew McAfee and Erik Brynjolfsson, “Big Data: The Management Revolution,” *Harvard Business Review*, XC (2012), 60–66. Schmidt’s figures have been disputed (<http://readwrite.com/2011/02/07/are-we-really-creating-as-much>), but the point is that the quantity of data is increasing exponentially.

The purpose of this Research Note is to show the extent to which the data revolution is consequential for the field of history. The surge in computing power and access to data-processing software and online resources have enabled historians during the past two decades to capture historical statistics on a much larger scale than before. The statistical records stored in the archives of the imperial powers, as well as in those of the former colonies, are now available for analysis on an unprecedented scale. The data revolution is especially valuable when applied to regions where written records are scarce, such as sub-Saharan Africa. This Research Note documents how a new generation of economists, geographers, and historians is rewriting African history using archival records (colonial sources written for purposes unrelated to current research questions) in conjunction with geographical, climatic, and demographic studies of the distant past to gather surprising new information.²

A NEW AFRICAN ECONOMIC HISTORY African history has always suffered from a limited number of written records. Not until the 1960s, when most African countries had gained independence, did historians in Africa and elsewhere begin to borrow methods from archaeology, linguistics, and oral history to investigate the continent's rich and varied past. Information about Africa's economic history at this point was in high demand. Africa's rapid economic growth rates during the late colonial and early postcolonial periods and the euphoria immediately following independence raised new questions about the effects of imperialism and colonialism and the possibilities of future prosperity. Economic-history departments were established in universities across the continent. Debates fostered within these newly established departments could be fierce, informed by the ideological battles between Marxist and liberal economics. Evidence in support of the various positions came most often from qualitative sources that originated during the colonial era, although the abundant statistical records were not completely neglected. Many pioneers of African data collection, such as Johnson, Szeresewski, Hopkins, Manning, Eltis, and Austen, began to assemble their data sets during the 1960s and 1970s.³

2 Patrick Manning, *Big Data in History* (Basingstoke, 2013).

3 Marion Johnson, "The Cowrie Currencies of West Africa Part I," *Journal of African History*, XI (1970), 17–49; *idem*, "Cloth as Money: The Cloth Strip Currencies of Africa," *Textile History*, XI (1980), 193–202; Robert Szeresewski, "The Process of Growth in Ghana, 1891–1911,"

By the 1980s, however, three forces had combined to push African economic history to the intellectual periphery. First, African economies were in decline. The oil shocks of the 1970s and the consequent debt crises and IMF (International Monetary Fund) structural-adjustment programs had diminished the resources available to African universities and deflated interest in explaining Africa's past fortunes. Second, as postmodernism became fashionable, historians concentrated more on cultural and social history than on economic development. Third, the cliometrics movement of the 1970s, which emphasized the mathematical and statistical aspects of economics, heightened the methodological barrier between economics and history. As a result of these trends, African economic history, both inside and outside Africa, went into a decline from which it did not begin to emerge until the 2000s. Led by prominent economists, the renaissance of African economic history coincided with the rise of Africa's economic prospects. A methodological and technological revolution within the economics profession certainly had something to do with the change. The new millennium also marked the dawn of the data revolution in African history.⁴

Acemoglu, Johnson, and Robinson are widely credited with the resurgence of African economic history through their provocative claim that disparities in economic development were largely due to the institutional legacies of European settlers. Their main contribution, however, was not so much their large new data set as it was the causal inference that they drew from it using a

Journal of Development Studies, I (1965), 123–141; Anthony G. Hopkins, “Economic Imperialism in West Africa: Lagos, 1880–1921,” *Economic History Review*, XXI (1968), 580–606; Manning, “Slaves, Palm Oil, and Political Power on the West African Coast,” *African Historical Studies*, II (1969), 279–288; David Eltis, “The Export of Slaves from Africa, 1821–1843,” *Journal of Economic History*, XXVII (1977), 409–433; Ralph A. Austen, “Slavery among Coastal Middlemen: The Duala of Cameroon,” in Suzanne Miers and Igor Kopytoff (eds.), *Slavery in Africa: Historical and Anthropological Perspectives* (Madison, 1977), 305–333.

4 Erik Green and Pius Nyambara, “The Internationalization of Economic History: Perspectives from the African Frontier,” *Economic History of Developing Regions*, XXX (2015), 68–78 (doi:10.1080/20780389.2015.1025744)—argue that African economic history declined only in Western institutions; Gareth Austin, “African Economic History in Africa,” *ibid.*, 79–94 (doi:10.1080/20780389.2015.1033686), suggests that it was also in decline in Africa. See also Morten Jerven, “African Growth Recurring: An Economic History Perspective on African Growth Episodes, 1690–2010,” *ibid.*, XXV (2010), 127–154; Austin and Stephen Broadberry, “Introduction: The Renaissance of African Economic History,” *Economic History Review*, LXVII (2014), 893–906; Fourie and Leigh Gardner, “The Internationalization of Economic History: A Puzzle,” *Economic History of Developing Regions*, XXIX (2014), 1–14.

novel econometric technique. Their claim about the persistence of institutions attracted the interest of a younger generation of economists, eager to tackle the issues of African poverty and underdevelopment. For instance, Nunn based his extensive study of the economic consequences of the slave trade on Curtin's Big Data project in African history from the 1960s, which culminated, following much debate and controversy, in Eltis and Richardson's Transatlantic Slave Trade Database. Although the various versions of the slave database that appeared during the past few decades have enriched a vast scholarship about this issue, Nunn's demonstration of a causal link between the number of slaves exported and the level of African GDP suggested a new direction for African economic history. Although the article later came under severe criticism, the consensus is that it rekindled economists' interest in African economic history.

Nunn's later work (with Wantchekon) identified mistrust in African societies as a legacy of the slave trade and (with Puga) showed how the slave trade pushed Africans into locations that still hamper their ability to trade. These contributions laid the foundation for a new trend in African economic history—the attempt to discover causal mechanisms in the past to explain conditions in the present.⁵

Given that few effects are more persistent than geographical ones, environmental information is crucial for linking historical events to current circumstances. Nunn and Puga posited rugged terrain as an explanatory variable. Alsan used temperature and humidity

5 Daron Acemoglu, Simon Johnson, and James A. Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, XCI (2001), 1369–1401; Nathan Nunn, "The Long-Term Effects of Africa's Slave Trades," *Quarterly Journal of Economics*, CXXIII (2008), 139–176; Philip D. Curtin, *The Atlantic Slave Trade: A Census* (Madison, 1972); Joseph E. Inikori, "Measuring the Atlantic Slave Trade: An Assessment of Curtin and Anstey," *Journal of African History*, XVII (1976), 197–223; Paul E. Lovejoy, "The Impact of the Atlantic Slave Trade on Africa: A Review of the Literature," *Journal of African History*, XXX (1989), 365–394; Eltis and David Richardson (eds.), *Extending the Frontiers: Essays on the New Transatlantic Slave Trade Database* (New Haven, 2008); Manning, *Slavery and African Life: Occidental, Oriental, and African Slave Trades* (New York, 1990); Lovejoy, *The African Diaspora: Revisionist Interpretations of Ethnicity, Culture and Religion under Slavery* (Boston, 1997); Ewout Frankema and Marlous van Waijenburg, "Structural Impediments to African Growth? New Evidence from Real Wages in British Africa, 1880–1965," *Journal of Economic History*, LXXII (2012), 895–926. The reason for the criticism is more explicit in the working paper version, available at http://eh.net/eha/wp-content/uploads/2013/11/Frankema_Waijenburg_o.pdf. Nunn and Leonard Wantchekon, "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, CI (2011), 3221–3252; Nunn and Diego Puga, "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, XCIV (2012), 20–36.

to assess the influence of the tsetse fly on development in Africa. Bhattacharyya advanced the prevalence of malaria as the most important factor in explaining African poverty, at least from a statistical perspective. Fenske and Kala's annual panel of African temperatures and slave exports indicated how past environmental shocks affected the slave trade. Papaioannou's research into court cases and the records of prisoners and homicides revealed how deviations from the long-term rainfall pattern increased conflict in colonial Nigeria. Papaioannou and de Haas extended this analysis to include other parts of Africa with similar results. Climatic and other environmental information geocoded at the micro-level enabled scholars to test causal persistence with a high degree of statistical accuracy.⁶

The new variables gleaned from the data revolution not only measure past events; they also serve effectively as contemporary outcome variables. In the absence of data about regional African economies, Michalopoulos and Papaioannou employed light density at night, obtained from satellite imaging, to test, on the one hand, the effect of precolonial ethnic institutions and, on the other, the effect of the borders created during the European powers' scramble for Africa on current economic performance. Notwithstanding the strong criticism that Michalopoulos and Papaioannou's data sources and estimation techniques have attracted, light density has become a popular outcome variable.⁷

6 Marcella Alsan, "The Effect of the TseTse Fly on African Development," *American Economic Review*, CV (2015), 382–410; Sambit Bhattacharyya, "Root Causes of African Underdevelopment," *Journal of African Economies*, XVIII (2009), 745–780; James Fenske and Namrata Kala, "Climate and the Slave Trade," *Journal of Development Economics*, CXII (2015), 19–32; Kostadis J. Papaioannou, "Climate Shocks and Conflict: Evidence from Colonial Nigeria," African Economic History Network Working Paper Series, 17 (2014); Papaioannou and Michiel de Haas, "Climate Shocks, Cash Crops and Resilience: Evidence from Colonial Tropical Africa," paper presented at the World Economic History Congress, August 3–8, 2015, Kyoto.

7 Stelios Michalopoulos and Papaioannou, "Pre Colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, LXXXI (2013), 113–152; *idem*, "Further Evidence on the Link between Pre-colonial Political Centralization and Comparative Economic Development in Africa," *Economics Letters*, CXXVI (2015), 57–62; *idem*, "National Institutions and Subnational Development in Africa," *Quarterly Journal of Economics*, CXXIX (2014) 151–213; Timothy Besley and Marta Reynal-Querol, "The Legacy of Historical Conflict: Evidence from Africa," *American Political Science Review*, CVIII (2014), 319–336; Nonso Obikili, "An Examination of Subnational Growth in Nigeria: 1999–2012," *South African Journal of Economics*, LXXXIII (2015), 335–357; Denis Cogneau and Yannick Dupraz, "Questionable Inference on the Power of Pre-Colonial Institutions in Africa," PSE Working Papers, No. 2014–25 (2014).

Contemporary survey data, now with geocoded observations, also frequently become outcome variables. A case in point is the “Afrobarometer survey” by Nunn and Wantchekon to measure levels of trust. Abel’s work regarding forced removals during apartheid also relies on the Afrobarometer survey. Michalopoulos, Putterman, and Weil’s use of Demographic and Health Survey (DHS) data in combination with information about the respondents’ ancestral ethnicity discovered that descendants of agriculturalists are wealthier and better-educated than the descendants of pastoralists.⁸

Not everyone agrees that such broad-brush studies, which link a historical episode to contemporary data, are always edifying. The main complaint about such an approach is that it “compresses history,” oversimplifying a more elaborate causality. In a recent book, Jerven criticizes what he calls the irresponsible manner in which economists have deployed econometrics to explain African development. In reviewing this book, de Waal calls for African economic history to be “liberated from the tyranny of econometricians.”⁹

African economic historians have proposed an alternative methodology that involves unearthing and digitizing colonial-era archival records. For the last decade, scholars have scanned colonial blue books (a compendium of records about the civil establishment, revenue and expenditure, and other statistical particulars of the colonies issued by the Colonial Office), tax censuses, voter rolls, marriage registers, etc., to study the population size, wages, incomes, education, fiscal systems, and transport networks of

8 The Afrobarometer survey series measures public attitudes about economic, political, and social matters in several sub-Saharan African countries. More information is available at <http://www.afrobarometer.org/>. Martin Abel, “Long-Run Effects of Forced Removal under Apartheid on Social Capital,” paper presented at African Economic History meetings, London School of Economics and Political Science, October 25–26, 2014; Michalopoulos, Louis Putterman, and David N. Weil, “The Influence of Ancestral Lifeways on Individual Economic Outcomes in Sub-Saharan Africa,” paper discussed at NBER Reporter Number 4: Program and Working Group Meetings (2014).

9 Austin, “The ‘Reversal of Fortune’ Thesis and the Compression of History: Perspectives from African and Comparative Economic History,” *Journal of International Development*, XX (2008), 996–1027; Jerven, *Africa: Why Economists Get it Wrong* (London, 2015); *idem* and Deborah Johnston, “Statistical Tragedy in Africa? Evaluating the Data Base for African Economic Development,” *Journal of Development Studies*, LI (2015), 111–115; Alex de Waal, “Liberating African Economic History from the Tyranny of Econometric,” review of Jerven, *Africa: Why Economists Get it Wrong*, available at: <http://africanarguments.org/2015/06/24/liberating-african-economic-history-from-the-tyranny-of-econometrics-by-alex-de-waal/>.

African societies. Rather than a lengthy survey of every contribution within this context, a single type of archival document—the military attestation form, which includes, among other things, soldiers’ heights—can represent the nature of the data revolution in African economic history. These individual-level observations produce a valuable proxy for living standards in the absence of other evidence.¹⁰

HOW HEIGHT DATA CAN HELP TO UNLOCK THE AFRICAN PAST The use of human heights or stature as a proxy for living standards is more than three decades old. Social scientists agree that height

10 Manning, “Historical Datasets on Africa and the African Atlantic,” *Journal of Comparative Economics*, XL (2012), 604–607; Frankema and Jerven, “Writing History Backwards or Side-ways: Towards a Consensus on African Population, 1850–2010,” *Economic History Review*, LXVII (2014), 907–931; Fourie and Erik Green, “The Missing People: Accounting for the Productivity of Indigenous Populations in Cape Colonial History,” *Journal of African History*, LVI (2015), 195–215; Frankema and van Waijenburg, “Structural Impediments”; Rönnbäck Klas, “Living Standards on the Pre-Colonial Gold Coast: A Quantitative Estimate of African Laborers’ Welfare Ratios,” *European Review of Economic History*, XVIII (2014), 185–202; Sophia Du Plessis and Stan du Plessis, “Happy in the Service of the Company: The Purchasing Power of VOC Salaries at the Cape in the 18th Century,” *Economic History of Developing Regions*, XXVII (2012), 125–149; Jerven, “A West African Experiment: Constructing a GDP Series for Colonial Ghana, 1891–1950,” *Economic History Review*, LXVII (2014), 964–992; Leandro Prados de la Escosura, “Output per Head in Pre-Independence Africa: Quantitative Conjectures,” *Economic History of Developing Regions*, XXVII (2012), 1–36; Fourie and Jan Luiten van Zanden, “GDP in the Dutch Cape Colony: The National Accounts of a Slave-Based Society,” *South African Journal of Economics*, LXXXI (2013), 467–490; Wantchekon, Marko Klačnja, and Natalija Novta, “Education and Human Capital Externalities: Evidence from Colonial Benin,” *Quarterly Journal of Economics*, CXXX (2015), 703–757; Fourie and Dieter von Fintel, “Settler Skills and Colonial Development: The Huguenot Wine Makers in Eighteenth Century Dutch South Africa,” *Economic History Review*, LXVII (2014), 932–963; Nonso, Obikili, “Social Capital and Human Capital in the Colonies: A Study of Cocoa Farmers in Western Nigeria,” *Economic History of Developing Regions*, XXX (2015), 1–22 (doi: 10.1080/20780389.2015.1012712); Jörg Baten, and Fourie, “Numeracy of Africans, Asians, and Europeans during the Early Modern Period: New Evidence from Cape Colony Court Registers,” *Economic History Review*, LXVIII (2015), 632–656; Gardner, *Taxing Colonial Africa: The Political Economy of British Imperialism* (New York, 2012); Fourie, Ada Jansen, and Krige Siebrits, “Public Finances and Private Company Rule: The Dutch Cape Colony (1652–1795),” *New Contree*, 68 (December 2013), 1–22; Frankema and van Waijenburg, “Metropolitan Blueprints of Colonial Taxation? Lessons from Fiscal Capacity Building in British and French Africa, c. 1880–1940,” *Journal of African History*, LV (2014), 371–400; Frankema, “Colonial Taxation and Government Spending in British Africa, 1880–1940: Maximizing Revenue or Minimizing Effort?” *Explorations in Economic History*, XLVIII (2011), 136–149; Papaioannou and Angus Edwin Dalrymple-Smith, “Political Instability and Discontinuity in Nigeria: The Pre-Colonial Past and Public Goods Provision under Colonial and Post-Colonial Political Orders,” *Economics of Peace and Security Journal*, X (2015), 40–53; Remi Jedwab and Alexander Moradi, “The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa,” *Review of Economics and Statistics* (forthcoming).

accurately reflects an individual's environmental conditions early in life, including access to nutrition and exposure to disease, and that changes in average height are reflective of changes in a society's living standards. Although heights might be expected to correlate positively with incomes, economic historians analyzing heights in Western Europe and North America discovered a surprisingly negative correlation between height and income during the early phase of industrialization in England and elsewhere. This apparent anomaly became known as the "Early Industrial Growth Puzzle" in Europe and the "Antebellum Puzzle" in the United States; for most of the last two decades of the twentieth-century, scholars on both sides of the Atlantic have attempted to explain it.¹¹

The height of nineteenth-century Africans was the subject of two articles by Eltis in 1982 and 1990, but the turn to heights to document the evolution in the living standards of African peoples during an era of unreliable data had to await Moradi and Baten's pioneering investigation of height differences between historical birth cohorts constructed from DHS data. But DHS data were not available until the colonial era ended in the 1950s. Another source was needed to investigate the colonial and the precolonial eras. Moradi found a sample of 1,046 Ghanaian recruits from World War I and World War II and 730 Kenyan recruits from World War II, adding information from surveys, to measure the effect of colonial policies on African living standards. In a follow-up work about the same question, his sample more than doubled.¹²

11 Robert W. Fogel et al., "Secular Changes in American and British Stature and Nutrition," *Journal of Interdisciplinary History*, XIV (1983), 445–481; *idem*, Stanley L. Engerman, and James Trussell, "Exploring the Uses of Data on Height: The Analysis of Long-Term Trends in Nutrition, Labor Welfare, and Labor Productivity," *Social Science History*, VI (1982), 401–421; Richard H. Steckel, "Height and Per Capita Income," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, XVI (1983), 1–7; Steckel, "Stature and the Standard of Living," *Journal of Economic Literature*, XXXIII (1995), 1903–1940; George Alter, "Height, Frailty, and the Standard of Living: Modeling the Effects of Diet and Disease on Declining Mortality and Increasing Height," *Population Studies*, LVIII (2004), 265–279; Angus Deaton, "Height, Health, and Inequality: The Distribution of Adult Heights in India," *American Economic Review*, XCVIII (2008), 468; John Komlos, "Shrinking in a Growing Economy? The Mystery of Physical Stature during the Industrial Revolution," *Journal of Economic History*, LVIII (1998), 779–802; *idem*, "Anomalies in Economic History: Toward a Resolution of the 'Antebellum Puzzle,'" *ibid.*, LVI (1996), 202–214; *idem* and Bjorn Alecke, "The Economics of Antebellum Slave Heights Reconsidered," *Journal of Interdisciplinary History*, XXVI (1996), 437–457.

12 Eltis, "Nutritional Trends in Africa and the Americas: Heights of Africans, 1819–1839," *Journal of Interdisciplinary History*, XII (1982), 453–475; *idem*, "Welfare Trends among the Yoruba

Moradi's groundbreaking work stimulated interest in military records elsewhere in Africa, often with startling results. Austin, Baten, and van Leeuwen found that in nineteenth-century West Africa, Ghanaian and Burkinabe recruits were notably shorter than northwestern Europeans but not shorter than southern Europeans. In twentieth-century West Africa, Cogneau and Rouanet found that the rate of increase in the heights of those born in Côte d'Ivoire and Ghana during the late colonial period, 1925 to 1960, was almost as fast as the rate observed in France and Great Britain during the period from 1875 to 1975. Inwood and Masakure found that colored South Africans were, on average, six cm shorter than white South Africans at the start of the twentieth century, a significantly smaller difference than today's eight cm. The findings from these studies now take their place within a larger literature about the living standards of indigenous populations across the world.¹³

in the Early Nineteenth Century: The Anthropometric Evidence," *Journal of Economic History*, L (1990), 521–540; Alexander Moradi and Baten, "Inequality in Sub-Saharan Africa: New Data and New Insights from Anthropometric Estimates," *World Development*, XXXIII (2005), 1233–1265. Other pioneering work included David E. Sahn and David C. Stifel, "Urban-Rural Inequality in Living Standards in Africa," *Journal of African Economies*, XII (2003), 564–597. Moradi, "Confronting Colonial Legacies: Lessons from Human Development in Ghana and Kenya, 1880–2000," *Journal of International Development*, XX (2008), 1107–1121, *idem*, "Towards an Objective Account of Nutrition and Health in Colonial Kenya: A Study of Stature in African Army Recruits and Civilians, 1880–1980," *Journal of Economic History*, LXIX (2009), 719–754.

13 Austin, Baten, and Bas Van Leeuwen, "The Biological Standard of Living in Early Nineteenth Century West Africa: New Anthropometric Evidence for Northern Ghana and Burkina Faso," *Economic History Review*, LXV (2012), 1280–1302; Denis Cogneau and Léa Rouanet, "Living Conditions in Côte d'Ivoire and Ghana, 1925–1985: What Do Survey Data on Height Stature Tell Us?" *Economic History of Developing Regions*, XXVI (2011), 55–82; Kris Inwood and Oliver Masakure, "Poverty and Physical Well-Being among the Coloured Population in South Africa," *Economic History of Developing Regions*, XXVIII (2013), 56–82; Joseph M. Prince and Steckel, "Nutritional Success on the Great Plains: Nineteenth-Century Equestrian Nomads," *Journal of Interdisciplinary History*, XXXIII (2003), 353–384; Howard Bodenhorn, "The Mulatto Advantage: The Biological Consequences of Complexion in Rural Antebellum Virginia," *ibid.* (2002), 21–46; Aravinda Meera Guntupalli and Baten, "The Development and Inequality of Heights in North, West, and East India 1915–1944," *Explorations in Economic History*, XLIII (2006), 578–608; Inwood, Les Oxley, and Evan Roberts, "Physical Growth and Ethnic Inequality in New Zealand Prisons, 1840–1975," *History of the Family*, XX (2015), 249–269 (doi:10.1080/1081602X.2015.1006653); Baten, Ines Pelger, and Linda Twrdek, "The Anthropometric History of Argentina, Brazil and Peru during the 19th and early 20th Century," *Economics & Human Biology*, VII (2009), 319–333; Baten, Mojgan Stegl, and Pierre van der Eng, "The Biological Standard of Living and Body Height in Colonial and Post-Colonial Indonesia, 1770–2000," *Journal of Bioeconomics*, XV (2013), 103–122; Baten and Matthias Blum, "Growing Tall but Unequal: New Findings and New Background Evidence on Anthropometric Welfare in 156 Countries, 1810–1989," *Economic History of Developing Regions*, XXVII (2012), S66–S85.

Information in attestations is important not only for measuring living standards over time but also for tracing the effects of different colonial policies. Cogneau and Moradi constructed a much larger sample, 11,940 recruits from Ghana and Togo, to evaluate how the partition of German Togoland after World War I influenced educational outcomes. The part of Togo that fell under French mandate had lower levels of literacy after the partition than did the parts that were ruled by the British. According to Cogneau and Moradi, the French authorities were hostile to missionary schools. Jedwab and Moradi's analysis of colonial railways *vis à vis* various economic outcomes, including heights, found that railways increased the heights of those born closest to them and generally improved income, boosting the production of cash crops like cocoa.¹⁴

Mariotti's investigation of the heights of black mineworkers in South Africa confirmed the positive effect of income improvement on height. A 1974 plane crash that killed seventy-three mineworkers provoked the Malawian government to ban the migration of mineworkers to South Africa, forcing South Africa to hire workers from the Transkei "homeland" in South Africa. Mariotti showed that the resulting improvement in household incomes for the newly recruited Transkeian mineworkers increased the heights of children born during or immediately after 1974, and only in those districts from which mineworkers came. Mariotti and Dinkelman later turned their attention to the effects of the plane crash and the sudden prevention of labor migration on Malawian households.¹⁵

Attestation records provide a snapshot of living standards in the absence of the other individual-level records that colonial authorities were often less-inclined to collect for indigenous populations than for Europeans. The demographic information that is available typically comes with a colonial-era bias that is difficult to exclude. Although height information collected for military purposes is

14 Cogneau and Moradi, "Borders that Divide: Education and Religion in Ghana and Togo since Colonial Times," *Journal of Economic History*, LXXIV (2014), 694–729; Jedwab and Moradi, "Permanent Effects."

15 Martine Mariotti, "Fathers' Employment and Sons' Stature: The Long-Run Effects of a Positive Regional Employment Shock in South Africa's Mining Industry," *Economic Development and Cultural Change*, LXIII (2015), 485–514; Taryn Dinkelman and Mariotti, "Does Labor Migration Affect Human Capital in the Long Run? Evidence from Malawi," unpub. paper (Dartmouth College, 2014).

unlikely to demonstrate such systematic bias, attestation forms can have a selection bias of their own. Not everyone qualified for military service. The minimum-height requirement meant that attestation heights were at times truncated to the left. Truncation is amenable to a statistical check (by using truncated regression models, for example), but another unobservable selection could always be lurking. Indeed, Bodenhorn, Guinnane, and Mroz have recently addressed this very possibility. They claim that the declining heights observed during the Industrial Revolution—the “Early Industrial Growth Puzzle”—reflects not a decline in living standards but a change in the military’s selection process. When returns in the private sector were small (at the start of the Industrial Revolution), many of the stronger (or taller) men probably chose the military as a career. As the economy grew, however, more of them would have returned to private-sector employment, with no concomitant increase in the military population, leaving the weaker (or shorter) men to join the army. Moreover, Inwood, Mariotti, and Fourie used data about South African recruits in World War I to show that this type of unobservable selection may also happen because of changes in military technology.¹⁶

Selection-bias concerns, however, cannot prevent heights from being a useful proxy for individual-level living standards in the absence of other measures. A comparison of the results derived from military attestation forms together with those derived from survey data, for example, suggests that selection into the military may not have been strong enough to undermine our confidence in earlier findings. The studies that have been done on heights exemplify the contribution of the data revolution in African history to new insights about how colonial-era policies affected the welfare of subjects who remained outside the remit of recordkeeping.

Besides heights, attestation forms offer a wealth of data that can contribute to fields outside economic history, in particular social history. The attestation forms of the South African Constabulary list several characteristics of each recruit, including the ability to speak an

16 See for example, Bodenhorn, Timothy W. Guinnane, and Thomas A. Mroz, “Sample-Selection Biases and the ‘Industrialization Puzzle,’” NBER Working Paper 21249 (2015); *idem*, “Caveat Lector: Sample Selection in Historical Heights and the Interpretation of Early Industrializing Economies,” NBER Working Paper 19955 (2014). Fourie, Inwood, and Mariotti, “Can Historical Changes in Military Technology Explain the Industrial Growth Puzzle?” unpub. paper (London School of Economics, 2014).

African language, the color of hair and eyes, vaccination history, the ability to ride a horse or swim, and the presence of any tattoos. A simple quantitative analysis can bring to light unsuspected details. More than one-in-five Boer recruits from the two Boer Republics had at least one tattoo, and the number of Boer recruits who were redheads was nearly twice that of the Scottish or Irish recruits.¹⁷

DATA TRANSCRIPTION, AUGMENTATION, AND COLLABORATION The use of large archival data sets is only one aspect of the data revolution. New tools for data transcription, augmentation, and collaboration allow for faster and less-expensive digitization, more rigorous investigation, and better replicability and comparability of results. The data revolution comprises not only new material but also a new method—the use of computational power and statistical techniques to expand the scale and scope of research questions.

Data transcription—the manual copying of archival sources, often into a digital format such as Excel—is laborious and expensive. As an example, Figure 1 shows the attestation of Diederick Alfred Joseph Yates, a Briton who enlisted in the South African Constabulary in 1902. Although the form is standardized and the text clearly legible to anyone familiar with late nineteenth-century handwriting, it does not allow for automated transcription. Optical character recognition (OCR) software is not yet able to capture the information contained in such an attestation with a high degree of reliability. Historians must therefore spend countless hours transcribing these attestations into analyzable text or train a research team to do it. Both options are costly. They can also cut corners, as Inwood and Masakure did when investigating South African soldiers who served in World War I and World War II. Although several hundred thousand attestations are available in the Defense Force archives, time and funding constraints forced Inwood and Masakure to limit their transcriptions to only 10,000 attestations for each of the wars. Such sampling introduces additional biases for which researchers must take account when analyzing the results. It also does not fully conform to the methodology now popular in the era of Big Data, which advocates the use of all available data.

17 Fourie, Inwood, and Mariotti, “‘Poor South Africa! Will No Nice English People Ever Come Out Here?’—The South African Constabulary of the Second South African War,” unpub. paper (Stellenbosch University, 2015).

Fig. 1 An Attestation Form from the South African Constabulary



There is reason to believe, however, that improved OCR software will soon make the transcription of archival sources, such as attestation forms, less costly. This rapidly developing technology—the most spectacular use of which is Google’s project to scan every book title in existence (estimated in 2010 to be about 130 million)—already has wide applicability in history, for example in the digitization of old newspapers. But printed texts are relatively easy for the OCR technology to digitize and transcribe. Eventually an OCR algorithm will be able to read a natural image—like the attestation in Figure 1—that contains written text. This technological breakthrough in computational linguistics could have far-reaching consequences for the field of economic history.¹⁸

18 Maya R. Gupta, Nathaniel P. Jacobson, and Eric K. Garcia, “OCR Binarization and Image Pre-Processing for Searching Historical Documents,” *Pattern Recognition*, XL (2007), 389–397; Sergey Milyaev et al., “Image Binarization for End-to-End Text Understanding in Natural Images,” paper presented at the Twelfth International Conference on Document Analysis and Recognition (ICDAR), August 25–28, 2013, Washington, D.C.

Data augmentation, the second step in the new methodology, signifies the addition of information from internal or external sources to the original data. Although supplementing existing data with new information has always been possible, researchers usually had to do it manually. Computerization and the development of matching algorithms have made this process much easier. In the case of the attestation forms, scholars can now match recruits' names and birth dates to, say, genealogical records. Not only can they learn more about their subjects; they can also solve previously intractable problems, such as calculating intergenerational mobility using height as an indicator of living standards. Data augmentation has countless new avenues to explore.¹⁹

The third step is collaborating across research networks. Data sharing has already resulted in several projects to compare historical measures. At a macroeconomic level, series such as the Penn World Tables and the Maddison project have provided scholars with cross-country comparisons in history. Manning's Big Data in History project aims to "create a world-historical archive that will trace the last four centuries of historical dynamics and change." The Global Collaboratory on the History of Labour Relations, based at the International Institute of Social History in Amsterdam, has assembled a team to collect quantitative benchmarks for a global labor history—including population size and composition, workforce size and composition, occupations, and types of labor—setting the years 1500, 1650, 1800, 1900, and 2000 as benchmarks (1950 is included for Africa). Van Zanden has undertaken the Clio Infra project, which aims to establish "a set of interconnected databases . . . containing worldwide data on social, economic, and institutional indicators for the past five centuries, with special attention to the past 200 years." This project has now evolved into the CLARIAH (Common Lab Research Infrastructure for the Arts and Humanities) project, consisting of a consortium of more than forty partners who store and share all types of data related to the arts and humanities in a standardized and user-friendly format.²⁰

19 See, for example, James J. Feigenbaum, "Automated Census Record Linking," unpub. paper (Harvard University, 2015), available at <http://scholar.harvard.edu/files/jfeigenbaum/files/feigenbaum-censuslink.pdf>.

20 Robert Feenstra, Robert Inklaar, and Marcel Timmer, "The Next Generation of the Penn World Table," NBER Working Paper 19255 (2013); Jutta Bolt and van Zanden, "The Maddison Project: Collaborative Research on Historical National Accounts," *Economic*

In African history, Frankema and his team have taken the lead in digitizing and transcribing the colonial blue books, facilitating research about colonial education, public finance, and population. In an award-winning study, Frankema and van Waijenburg demonstrated that African real wages were significantly above subsistence levels and rising for most of the colonial period. In some places, wages were much higher than comparable figures for Asia. That study suggests that the idea of sub-Saharan Africa as always having been poor is not supported by the evidence.²¹

Wages and prices can reveal much more than simply comparative levels of development. Real wages also help to examine changes in the degree of inequality between the rich and the poor (or between the colonizer and the colonized), a topic that has attracted much attention following the publication of Piketty's *Capital in the Twenty-First Century*. Because historical wage and price series are often not readily available, several scholars are turning to such innovative sources as slave-ship provisions, Dutch East India Company records, or agricultural journals.²²

On a broader scale, Jerven has tracked macroeconomic trends backward into the past. By reconstructing African GDP measures, he has found considerable discrepancies between World Bank, IMF, and Penn World Table estimates. In two highly acclaimed books, Jerven emphasizes that Africa's image as the "hopeless

History Review, LXVII (2014), 627–651; Manning, *Big Data*. For the Global Collaboratory, see <https://collab.iisg.nl/web/labourrelations>. Van Zanden et al. (eds.), *How Was Life? Global Well-Being since 1820* (Paris, 2014).

21 Frankema and Jerven, "Writing History Backwards or Sideways"; Frankema, "The Origins of Formal Education in Sub-Saharan Africa: Was British Rule More Benign?" *European Review of Economic History*, XVI (2012), 335–355; *idem*, "Colonial Taxation and Government Spending in British Africa, 1880–1940: Maximizing Revenue or Minimizing Effort?" *Explorations in Economic History*, XLVIII (2011), 136–149. *Idem* and van Waijenburg, "Structural Impediments," won the 2013 Arthur H. Cole prize for the best paper published in the *Journal of Economic History*.

22 Thomas Piketty (trans. Arthur Goldhammer), *Capital in the Twenty-First Century* (Cambridge, Mass., 2014); *idem*, and Emmanuel Saez, "Inequality in the Long Run," *Science*, CCCXLIV (2014), 838–843. For innovative sources, see, for example, Inikori's attempt to extract African market-price data from the slave-trade records to measure precolonial African economies at <https://www.wilsoncenter.org/person/joseph-inikori>; Dalrymple-Smith and Frankema's collection of prices from slave-ship provisions at http://www.wageningenur.nl/upload_mm/7/8/8/aa9b6927-3229-4323-b6f4-0a343dea34b4_Dalrymple-Smith%20Frankema%20-%20Paper.pdf; Sophia du Plessis and Stan du Plessis, "Happy in the Service of the Company"; Willem Boshoff and Fourie, "When did Globalization Begin in South Africa?" Stellenbosch Working Paper Series WP10/2015.

continent”—famously encapsulated by the May 13, 2000, cover of *The Economist*—is largely a construction of the 1980s and 1990s. Bolt and Gardner are compiling fiscal revenue and expenditure data at the local level for the late colonial period. They pair these new records with evidence about precolonial states to assess the extent to which the fiscal capacity of local-government units reflects precolonial state centralization. Moreover, in an attempt to understand the evolution of inequality during the colonial era (in the absence of information about income at the individual level), Bolt and Hillbom are compiling colonial-era records of occupations and labor-market structures.²³

These reinterpretations of the African past are the result of digitizing and transcribing the vast amounts of data available in colonial archives—projects that are ongoing. Meier zu Selhausen is digitizing and transcribing vast quantities of demographic records kept in missionary-station archives. These records are not without problems. Selection into formal Christian mission stations may carry hidden biases, and gaining the trust of bishops and others anxious that the data should be put to proper use is often as time-consuming for researchers as is the analysis. But the effort is worth the cost because these records provide a glimpse of African demographic changes and living standards unrecorded in the colonial blue books.²⁴

The records available for the European settlers and their activities are more detailed. Green, von Fintel, and Fourie are constructing an annual panel data set of several thousand settler farmers for more than 140 years. Once complete, this data set will offer a wide-ranging and informative account of eighteenth- and nineteenth-century life in colonial South Africa and allow economists to investigate the

23 Jerven, “African Growth Recurring”; *idem*, “For Richer, For Poorer: GDP Revisions and Africa’s Statistical Tragedy,” *African Affairs*, CXII (2012), 138–147; *idem*, *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It* (Ithaca, 2013); *idem*, *Africa: Why Economists Get It Wrong*; Bolt and Gardner, “De-compressing History? Pre-colonial Institutions and Local Government Finance in British Colonial Africa,” paper presented at the Economic History Association meeting, September 11, 2015, Nashville; Bolt and Ellen Hillbom, “Potential for Diversification? The Role of the Formal Sector in Bechuanaland Protectorate’s Economy, 1900–65,” *Economic History of Developing Regions*, XXX (2015), 1–30.

24 Felix Meier zu Selhausen, “Missionaries and Female Empowerment in Colonial Uganda: New Evidence from Protestant Marriage Registers, 1880–1945,” *Economic History of Developing Regions*, XXIX (2014), 74–112; *idem* and Jacob Weisdorf, “A Colonial Legacy of African Gender Inequality? Evidence from Christian Kampala, 1895–2011,” *Economic History Review*, LXIX (2016), 229–257.

economic consequences of such events as the abolition of slavery and smallpox epidemics. Even though these statistical records were collected for the European colonial administration, they contain detailed demographic information about the Khoisan, a population group almost completely neglected in other sources. The scale of this project—several thousand individuals for each year—does not allow for matching individuals manually across years. Instead, an algorithm confers unique individuals with a distinct ID that permits them to be observed throughout their entire lifetime. By adding genealogical records, we can also expand the algorithm to link families across generations. Such an intergenerational panel would not have been feasible without the advent of enhanced computing and powerful statistical techniques.²⁵

THE DIFFUSION OF THE DATA REVOLUTION Has the data revolution produced a better understanding of the African past and the manner in which historical factors still shape African destinies? It may be too early to pass judgment. However, the new approach has generated interest far beyond the narrowly defined disciplinary borders of economic history. Econometricians, geographers, evolutionary biologists, linguists, demographers, sociologists, and computer scientists are beginning to exploit the rich quantitative history of Africa using their own methods. Economic historians should welcome this trend, not only because it generates an audience for their work but also because it provides new ways to test conjectures and hypotheses. The risk, however, is that some economic historians may be left behind, anchored to methods created at a time when quantitative data were less available. Many of the scholars who lack the statistical skills that the data revolution requires are based in Africa, geographically isolated from the interdisciplinary opportunities to acquire them. Of the more than fifty authors cited in the section above on the new economic history of Africa, fewer than ten are from Africa.²⁶

25 Fourie and Green, “Missing People”; Baten and Fourie, “Numeracy of Africans”; Jeanne Cilliers and Fourie, “New Estimates of Settler Life Span and Other Demographic Trends in South Africa, 1652–1948,” *Economic History of Developing Regions*, XXVII (2012), 61–86.

26 For evidence of Africans’ lack of the technical skills to keep up with advances in their own economic history, see the list of contributors to the special issue “The Economics of Apartheid,” published in *Economic History of Developing Regions*, XXIX (2014). For the introduction, see

Addressing this dearth of African scholars who are equipped to participate in the field is necessary for reasons of both fairness and intellectual diversity. As the new African economic history has demonstrated, the negative effects of colonialism still linger into the present; redressing these injustices by helping Africans to engage in, and contribute to, their own history is the first step to a more just intellectual discourse. But drawing African scholars into the field is likely to create more than just equality. It will also encourage the development of new questions, sources, approaches, and answers that will broaden and deepen the scope of African economic history.

A debate in a recent issue of *Economic History of Developing Regions* summarizes the factors that contribute to the current intellectual segregation. Green and Nyambara averred that economic history research at African universities “is not only strong, but remained vibrant even when African economic history was on the decline at universities elsewhere. The lack of visible output in major economic history journals is thus not a sign of weakness. Instead it is an effect of the increased methodological specialization of economic history in the Western world.” Green and Nyambara thus promote the engagement of Western economic historians with African scholars to prevent “regional isolation. In response, Austin suggested that “resource constraints,” along with “institutional constraints and intellectual priorities,” are responsible for the slow adoption of quantitative techniques by African scholars: “The overwhelming priority that economics departments in Africa rightly give to the study of current problems does not seem to be combined with an awareness of the uses of history in fulfilling this mission. Meanwhile, the institutionalization of the humanities/social sciences divide in many universities has made it less likely that history graduates will be equipped to combine qualitative and quantitative techniques, let alone focus on the latter.”²⁷

Although online access may bring African students into closer contact with new methodological tools, the existing disparities

Mariotti and Fourie, “The Economics of Apartheid: An Introduction,” *Economic History of Developing Regions*, XXIX (2014), 113–125. Gardner and Fourie discuss this problem for other developing regions in “The Internationalization of Economic History: A Puzzle,” *Economic History of Developing Regions*, XXIX (2014), 1–14.

27 Green and Nyambara, “Internationalization of Economic History,” 68; Austin, “African Economic History in Africa,” 91.

may well become more entrenched. It is not entirely clear which incentives will minimize the likelihood of a continued isolation of African scholars, or who will offer them. One approach endorsed by Green and Nyambara, as well as by Austin, is to encourage collaborative work through joint funding applications and research programs. The danger of this strategy, however, is that the non-African partners will tend to drive the research agendas. Another, more promising, approach is for European and U.S. universities to recruit and subsidize Ph.D. students from Africa who could begin the slow but sustainable process of training their African colleagues and students in the new methods when they return.²⁸

African economic history has already gained much from the data revolution of the past two decades. Although the major historiographical debates persist, we now know, for example, that the slave trade has made Africans wary of both their neighbors and their political leaders, that real wages during the early twentieth century were higher in many African countries than in Asian countries, and that colonial railways boosted production and altered the spatial distribution of cities. The continuing projects to transcribe and digitize large numbers of colonial and postcolonial records are likely to advance our knowledge about Africa's economic past.²⁹

African economic history is, however, not just about the past. A major issue for economic historians concerns how to explain Africa's apparent failure to sustain growth and development. The renaissance of African economic history at the start of the twenty-first century has coincided with a rise in African economies. But for a variety of external and internal reasons, the momentum seems to be slowing as pessimism about Africa's prospects increases. Can the tools of the data revolution bring a new understanding to Africa's past and promote policies for a better future? The answer is likely to be negative if the African beneficiaries of this would-be prosperous future cannot take part in the conversation and if

28 An ongoing project to write a freely downloadable textbook about African economic history recognizes that technology can leapfrog many of the traditional constraints to accessing higher education. See <http://www.aehnetwork.org/textbook/>.

29 Nunn and Wantchekon, "Slave Trade and the Origins of Mistrust"; Frankema and van Waijenburg, "Structural Impediments"; Jedwab and Moradi, "Permanent Effects."

non-African scholars in Europe or the United States fail to establish long-term partnerships with African universities and their faculties.

“The past is never dead. It isn’t even past,” William Faulkner once said. This statement is nowhere more apt than in Africa, where the deleterious effects of colonialism and slavery persist. Equipping African scholars with the tools to take part in the data revolution is not only necessary to redress the inequalities of the past but also vital to build a thriving interdisciplinary academic discourse.