Research Article

Engagement in the Knowledge Economy: Regional Patterns of Content Creation with a Focus on Sub-Saharan Africa

Sanna Ojanperä

University of Oxford, UK

Mark Graham

University of Oxford and The Alan Turing Institute, UK

Ralph K. Straumann

University of Oxford, UK

Stefano De Sabbata

University of Leicester, UK

Matthew Zook

University of Kentucky, USA

Abstract

Increasing digital connectivity has sparked many hopes for the democratization of information and knowledge production in sub-Saharan Africa. To investigate the patterns of knowledge creation in the region compared to other world regions, we examine three key metrics: spatial distributions of academic articles (traditional knowledge production), collaborative software development, and Internet domain registrations (digitally mediated knowledge production). We find that, contrary to the expectation that digital content is more evenly geographically distributed than academic articles, the global and regional patterns of collaborative coding and domain registrations are more uneven than those of academic articles. Despite hopes for democratization afforded by the information revolution, sub-Saharan Africa produces a smaller share of digital content than academic articles. Our results suggest the factors often framed as catalysts in the transformation into a knowledge economy do not relate to the three metrics uniformly. While connectivity is an important enabler of digital content creation, it seems to be only a necessary, not a sufficient, condition; wealth, innovation capacity, and public spending on education are also important factors.

> In times past, we searched for gold, precious stones, minerals, and ore. Today, it is knowledge that makes us rich and access to information is all-powerful in enabling individual and collective success. (Lesotho Ministry of Communications, Science and Technology, 2005, p. 12)

To cite this article: Ojanperä, S., Graham, M., Straumann, R. K., De Sabbata, S., & Zook, M. (2017). Engagement in the knowledge economy: Regional patterns of content creation with a focus on sub-Saharan Africa. *Information Technologies & International Development, 13,* 33–51.

ENGAGEMENT IN THE KNOWLEDGE ECONOMY

The knowledge economy[1] is a concept that has captured the imagination and spawned a wealth of literature in academic, business, and policy fields.[2] That concept, and the related ideas of an information economy, information society, and network economy all broadly focus on the increasing importance of codified information and knowledge for engaging and capturing value in an increasingly globalized and digitalized economy.[3]

The varied discourses on the knowledge economy postulate the role of knowledge as an economic output in itself and as an input that strengthens economic processes (Lundvall & Johnson, 1994), frequently employing technology (and information and communication technologies [ICTs] in particular). The evidence on the positive economic impacts of knowledge-enhanced products and processes often derives from high- or middle-income countries. Developing countries tend to be described as planning or embarking on a journey of transformation into knowledge economies and awaiting the economic gains that such a step is expected to lead to (Murenzi & Hughes, 2006; Udo & Edoho, 2000).

A few indices that attempt to measure knowledge economies have been published in recent years. While some of the indices focus on Europe or on specific developed countries, the World Bank Knowledge Economy Index (KEI, 2012) aims to highlight the global differences in the embeddedness of knowledge in economic activity. Until its last year of publication in 2012 the KEI painted a very uneven picture of the global knowledge economy, where the highly developed countries topped the rankings, and developing countries occupied the bottom of the index, with sub-Saharan African (SSA) countries ranking the lowest. Regarding the position of developing countries in the knowledge economy, Afele describes the "painful state of poverty and the underlying knowledge and network deficits among communities of the developing world" (2002, p. 7).

Yet, many see the potential for change, especially for SSA, where growing telecommunications revenues have recently prompted a surge of investment in telecom infrastructure (Gaibi, Maske, & Moraje, 2010). The growth of the software industry in certain emerging economies over the last few decades (Arora & Gambardella, 2004) has further fueled the continent's hopes of IT-led development. This has led a few countries such as South Africa and Rwanda to adopt a strong focus on IT in their plans for economic development (Casado-Lumbreras, Colomo-Palacios, Ogwueleka, & Misra, 2014; Lacity, Willcocks, & Rottman, 2008). A few studies that focus specifically on Africa indicate that the continent is starting to integrate with, or may be on its way to transform into, a knowledge economy. Powell, for instance, tellingly notes, "The concept of digital divide has tended to dominate the relatively sparse analysis of what the 'Information Age' heralds for Africa" (2001, p. 242). Only a few thematic studies have been carried out on country level, for example, in South Africa (Blankley & Booyens, 2010; Waghid, 2002) and Rwanda (Murenzi & Hughes, 2006). Many of the articles balance documentation of the digital divide and its hindering effect on the transformation into a knowledge economy with a positive outlook on the changes that integrating information and knowledge into economic activity could provide for the continent (Britz, Lor, Coetzee, & Bester, 2006; Juma & Agwara, 2006). Consequently, it seems that in their ICT policies, African governments echo technologically deterministic views and are eager to transform their economies to feature more knowledge-intensive processes and products. This follows a logic reminiscent of early modernization thinking in which we could expect many benefits such as increased growth and productivity in the telecom sector and other key sectors such as agriculture, banking, consumer goods, and extractives as well as reduced inequality (Friederici, Ojanperä, & Graham, 2017). However, Zegeye and Vambe (2006) caution that knowledge production and publishing in Africa has traditionally been (and continues to be) dominated by Western experts. Some of the academic literature and ICT policies underline the importance of creating local or indigenous information and knowledge, but few of these studies offer an analysis of knowledge creation within African countries (Powell, 2001; Zegeye & Vambe, 2006).

1. In this article, we occasionally refer to the knowledge economy not as a way to argue that a singular knowledge economy exists, but, rather, to attempt to refer to the constellation of economic activities and characteristics that make up many people's and organizations' conceptions of the term.

2. For further reading, see Powell and Snellman (2004) on the knowledge economy, Sörlin and Vessuri (2006) on the knowledge society, Castells (2003) on the information economy, Carmody (2013) on the information society, and Lundvall and Johnson (1994) on the learning economy.

3. When addressing information and knowledge, we refer to Kitchin's (2014) conceptualizations of information as linked elements of data, and knowledge as organized information.

The benefits of knowledge economies are generally acknowledged as well as the fact that low-income countries, many of them located in SSA, have largely not yet transformed into knowledge economies. However, the role that geography plays in this phenomenon deserves a closer look. Early thinkers on the topic such as O'Brien (1992) suggest that as a result of increasing connectivity, and the access it provides to new economic opportunities, the role of geography might be diminishing globally. Focusing on the increased global connectivity enabled by modern ICTs, Powell describes how the global network economy enables anyone with a commercially exploitable idea and appropriate links to engage with it, "if not on a basis of equality, at least as a player from anywhere on earth" (Powell, 2001, p. 247).

On the contrary, some state that geography matters more than ever as many forms of valuable knowledge are difficult to transfer, and knowledge-intensive economic activity tends to cluster in certain locations (Markusen, 1996; Sonn & Storper, 2008). Markusen describes advances in transportation and information as leading to "slippery" production spaces where the production of many goods and services is transnationalized. She finds that in the "slippery space" there are "sticky places," where knowledge-intensive activities tend to be concentrated (Markusen, 1996).

As the concept of the knowledge economy is notoriously difficult to measure and compare across countries, previous work on the geographies of codified knowledge has focused on particular aspects or segments of it, for instance, patents (Balland & Rigby, 2015; Jaffe & Trajtenberg, 2002; Sonn & Storper, 2008), citations (Jaffe & Trajtenberg, 2002), and innovation systems (Cooke, 2001). Despite the centrality of digital connectivity to the knowledge economy discourse, studies of the current geographies of digital knowledge and information on online platforms are rare. A recent article demonstrates that in addition to the geographies of uneven access to contemporary modes of communication, uneven geographies of participation and representation are also evident on platforms that mediate, host, and deliver different types of geographic information (Graham, De Sabbata, & Zook, 2015). Studying Wikipedia, Graham, Hogan, Straumann, and Medhat (2014) have established that the geographies of its content are highly uneven and denote the societal attitudes toward learning, information sharing, and broadband Internet accessibility as some of the contextual factors that constrain digital information production.

Given that the role of geography might diminish or increase the distribution of codified knowledge, it is surprising that the knowledge economy discourse in SSA rarely features studies of the particular geographies of digital content, even despite its focus on the role of telecommunications. This content involves measuring user activity and participation online and has the potential to capture patterns of information- and knowledge-intensive activities, which are otherwise difficult to measure in (near) real time. We thus argue for a need to better understand how access to and participation in constructing information and knowledge vary across space and across a wider variety of digital avenues. While access to digital means of communication such as mobile phones and the Internet is a useful metric, this access is loosely tied to knowledge production. Instead, we argue that digitally mediated participation in information- and knowledge-intensive activities offers a metric that more closely measures human capacity and skills. Thus, we contend it is essential to complement the knowledge economy discussion, chiefly dominated by studies of academic publications, patents, citations, and innovation systems, with a study of digital content that reflects user participation. An analysis of digitally mediated traces of skills and information might thus offer a way to better detect the boundaries of contemporary knowledge economies. This leads us to pose our first set of research questions:

RQ1a: What are the geographies of content development within SSA?

RQ1b: How does the production of content in SSA compare to other world regions?

To address the gap in research on digital content, we examine the geography of activities in collaborative software development (using the GitHub[4] platform) and the registration of top-level domains. While there are other indicators we could have included in the analysis, we selected these two because they have a global reach and they measure two distinct, but important, segments of the knowledge economy. Using the collaborative coding metric from GitHub provides a meaningful and global indicator of programming skills and

4. GitHub is a code-hosting platform, through which users can develop software and share code.

Table 1. Descriptive Statistics of the Dependent Variables (logarithm transformed).

Dependent Variable	Mean	Median	Standard Deviation	Year	# of Country Obs.	# of Units Across All Countries
Academic articles (l)	6.38	6.11	2.61	2011	168	1,799,695
GitHub commits (l)	8.50	8.57	3.48	2013	171	35,069,679
Domain registrations (l)	10.69	10.61	3.00	2013	172	245,141,725

Note: As the dependent variables had skewed heavy right tails, we logarithmized them to reduce the skewness and allow the patterns in the data to emerge more clearly in the analysis.

activity—a key skill within knowledge-based activities—and because it is open source, the barriers of participation (and bias) are lower than for other indicators. Domain name registration provides a measure of the volume of content and knowledge placed online. The global institutional and marketing structures of domain registration equalize the ability to set up and use Internet domains. We then contrast these distributions with an analysis of the geography of the production of academic articles as a traditional form of knowledge production. We chose this indicator because we consider academic publishing to offer a relatively uniform measure of knowledge-intensive output and because it consistently reflects global geographies of knowledge since the process of peer-reviewed publishing and the resulting construction of the scientific record are similar worldwide. Because of Africa's low levels of knowledge production and the hopes that have arisen that greater access to ICTs might bring about a knowledge revolution measurable via these alternative metrics, this article examines whether greater access to digital tools truly has allowed the region to transcend some of its traditional constraints.

As outlined above, the growth of knowledge economies is often attributed to a set of drivers such as education, human capital, innovation, and connectivity (Kolo, 2009), yet few studies measure the relationship between these drivers and the knowledge economy or segments of it. In African policies and organizational reports, the role of ICTs is often highlighted as the enabling and driving force, while donor organizations tend to look for a more balanced approach (Marker, McNamara, & Wallace, 2002; McNamara, 2003). African ICT strategies seem to prioritize the development of connectivity and ICT infrastructure and tend to frame their policies around those drivers. For instance, the Ghanaian government states:

> With the emerging information age, ICTs are seen by a number of countries as critical for achieving progress in economic and social development. These technologies are offering developing countries like Ghana a window of opportunity to leap-frog the key stages of industrialization and transform their subsistence agriculture-dominated economies into service-sector driven, high value-added information and knowledge economies that can successfully compete on the global market. (Republic of Ghana, 2003, p. 13)

Therefore, we pose a second research question:

> *RQ2: To what extent do the drivers commonly associated with knowledge economies (e.g., GDP, broadband Internet connections, education, and innovation) explain these geographies, and are there significant differences in how these factors explain the variance in traditional and digital content?*

Data: Dependent Variables

Table 1 presents the summary statistics of the three dependent variables.

Academic Articles

To approximate the level of domestic activities in formal research content creation, we use annual data on the number of academic articles published per country. Scientific output in the form of academic articles represents the progress of science. Publication of academic articles and the permanent scientific record they form are central for the codification of knowledge and a key enabler of knowledge-intensive processes. Beyond being an indicator often included in knowledge economy indices, we believe that academic articles offer a relatively uniform measure of knowledge-intensive output, as the process of peer-reviewed publishing and the way in which it constructs a permanent scientific record are similar worldwide. In contrast, other systems for knowledge-intensive outputs such as registering patents and innovation systems are known to vary greatly among countries and regions (Griliches, 1990). We measure the relative achievements of countries in producing academic articles through author affiliation locations in the Web of Science citation database, produced by Thomson Reuters. The most recently available data (from 2011) were chosen for analysis.

Web of Science has been noted as the world's largest accessible citation database. We chose it for our analysis due to this wide-ranging coverage. However, as with any database, it has its limitations. The database does not cover all published and peer-reviewed journals, and some disciplines are represented more fully than others. Further, the scholarly productivity from developing regions, and the African continent in particular, is hindered by challenges related to technology, sociopolitical factors, environmental and economic factors, and changing trends (Lauf, 2005; Nyamnjoh, 2004; Ondari-Okemwa, 2007). Yet, because Web of Science is often used as a benchmarking tool (i.e., journals are often ranked based on their Web of Science impact factor), it remains an important source of data for drawing an international comparison of the level of formal research content creation.

Collaborative Coding

We measure activities in collaborative software development using data from GitHub. GitHub is a code-hosting platform, through which users can develop software and share code. At the time of this writing, GitHub reported 11.9 million users working on 30 million projects, making it arguably the currently most widely used platform for collaborative software development (GitHub, 2015). Analyzing collaborative coding offers a way to measure programming skills and activities that are central to various information- and knowledge-rich activities such as software development. Other knowledge economy activities also depend on skilled programmers, and the open-source nature of GitHub makes it an easily adoptable application and thus a good indicator of coding activity in many countries.

GitHub enables researchers to analyze users and platform activity through an application programming interface (API) and data downloads. Using this data, we operationalize activities in collaborative software development by measuring the number of GitHub commits. *Commits* are individual units of content contribution to GitHub. They typically encompass changes to the code of a software project, but can also feature changes to documentation or supporting artwork such as graphical user interface components. We opted to use GitHub data from 2013 in this study. While GitHub commit data can be downloaded from 2011 through 2014, at the time of this writing, data on the majority of the explanatory variables are available only up to 2013. The GitHub commit data are not automatically associated with a country, but we geocoded every commit where the committing user had a user location attribute that could be parsed, using the Edina Unlock Places geocoding API for country name lookups.

While GitHub is the largest collaborative software development service, these data have some limitations. Only about 25% of users choose to indicate their location, and those users account for about 45% of commits. However, we have no reason to suspect that the tendency to release this information has any significant geographic biases. Furthermore, we acknowledge that the process of geocoding itself is subject to a margin of error. In a recent study, Lima, Rossi, and Musolesi (2014) assessed the quality of their geocoding of GitHub location information and found the rate of correctly geocoded profiles at about 90%. While we cannot verify the geocoding through cross-validation, we manually reviewed all toponyms attached to more than 10,000 commits in our data set and top toponyms per country. We therefore believe that our geocoding results are a good estimate of the underlying spatial distribution. While other approaches to measuring software production on GitHub exist, such as the number of projects (Kalliamvakou et al., 2014), users, or lines of code written, it has been argued that a commit more adequately captures the full range of potential contributions (Adams, Capiluppi, & Boldyreff, 2009).[5] Finally, GitHub faces competing services globally (for instance, CSDN and GitCafe in China) that may limit local user adoption. Additionally, recent research has reported the tendency of groups of cooperating developers to cluster in certain regions (Takhteyev & Hilts, 2010). Nonetheless, it

5. It is important to note that because of these data measure commits, it is possible we might see different levels of per-capita user engagement in different places.

remains that GitHub is a globally dominant platform at the time of our analysis. For all these considerations, we regard GitHub commits to be an important, if imperfect, proxy for digital knowledge creation (and likely the best available metric in this area).

Domain Registrations

Because domain names are a fundamental part of the Internet's infrastructure, they are a pervasive and important indicator of the number of entities providing online information and knowledge. As actors and activities in the knowledge economy rely on modern ICTs, the amount of accessible electronic content—as approximated by domain registrations—offers a perspective into countries' online presences (or absences). In analyzing the number of domain registrations associated with different countries, we total the three types of top-level domains (TLDs): generic TLDs (gTLDs) such as ".com" or ".net," country-code TLDs (ccTLDs) such as ".cn" for China, and internationalized TLDs (IDNccTLDs) that are ccTLDs in non-Latin characters such as Arabic. We decided to use data from 2013 to keep the two digital content creation variables on the same timescale.

To associate the TLD data with countries of registration, each type of TLD requires its own process. For gTLDs, the WHOIS record (a record of the address details of the person or organization registering a domain name) is geocoded, and each domain is assigned to the respective country. On the other hand, ccTLDs and IDNccTLDs are assigned to their respective countries (e.g., all ".cn" domains are assigned to China), as previous research (Zook, 2001) categorizes them as emblematic of local content production.

Some countries feature ccTLDs that have become widely popular because they are meaningful abbreviations (such as Tuvalu's ".tv" domain for the entertainment industry), form a word in a certain language (such as ".me"), form a word with the domain name (such as "bur.st"), or because they evoke an idea (such as ".io," which is used for start-ups). To detect upward-biased national ccTLD numbers, we compared the number of ccTLDs with Internet penetration. If the ccTLD number was very high in countries with low Internet penetration, we sought further information about the use of the ccTLD and discarded countries whose ccTLD seemed to be inflated by outside registrations (see Table 2 for omitted countries).

Because the remaining TLDs have a significant link to the address of the person or organization that registered them, we consider the TLD dataset to be an important proxy of digital online content creation.

Data: Explanatory Variables

To operationalize the explanatory variables for RQ2, we use data from the World Bank's World Development Indicators. To measure a country's innovation capacity, we use data from the World Economic Forum's Global Competitiveness dataset. We additionally consider two variables, which we expect to control for differences between countries and better evaluate the research questions, given that the data are not experimental. We include expenditure on education as a percentage of GDP in all the models and international trade as a percentage of GDP in the models with domain registrations as a dependent variable because we expect domain registrations to be more closely linked to trade.

When data for explanatory variables are unavailable for the same year as the dependent variable, we use the closest available year, seeing that the data for most variables of interest are steady over recent years for most countries. As many of the variables have skewed distributions and as some of them are pareto-distributed, we use the logarithmic forms for those that seem to benefit from the transformation. Table 3 presents the summary statistics of the explanatory variables.

Given the limited availability of data, the analysis of academic articles is modeled on data from 2011 and earlier, while the analyses of collaborative coding and domain registrations are modeled on data from 2013 and earlier. We believe the difference in the years is not large enough to produce sizable discrepancies in the way our variables behave between the two periods. We pay special attention to the different periods during our analysis, but we do not expect them to significantly impact our analysis and findings.

We acknowledge that while academic articles reflect a domain of knowledge with long and established traditions, the two digitally mediated types of content have shorter histories. Domain names have been registered actively since the early 1990s, but its 2008 launch makes GitHub a relatively young platform. Web of Knowledge and GitHub are leaders in their respective fields, but neither can account for the full volume of academic

Table 2. Countries Omitted Due to Inflated Domain Registrations.

Country	ccTLD	Reason for Omission
Armenia	.am	Used by the media industry
Federated States of Micronesia	.fm	Used by the media industry
Tuvalu	.tv	Used by the media industry
Mauritius	.mu	Used by the music industry
Ascension Island	.ac	Used by education-related websites
Réunion	.re	Used by real estate agents
Samoa	.ws	Used as an abbreviation for *website*
Montenegro	.me	Used for personal websites
Cocos Islands	.cc	Used as an alternative to .com
Cameroon	.cm	Used as an alternative to .com to exploit typing errors
Niue	.nu	Means *now* in Danish, Dutch, and Swedish
American Samoa	.as	The suffixes *AS* and *A/S* are used in some countries for joint stock companies
British Indian Ocean Territory	.io	Used by start-up companies
Sâo Tomé and Príncipe	.st	Used worldwide in several ways
Central African Republic	.cf	Can be registered free of charge
Gabon	.ga	Can be registered free of charge
Mali	.ml	Can be registered free of charge
Tokelau	.tk	Can be registered free of charge

Table 3. Descriptive Statistics of the Explanatory Variables.

Explanatory Variable	Mean	Median	Standard Deviation	# of Country Obs.
Fixed broadband connections 2011 (l)	12.04	12.05	2.96	168
Fixed broadband connections 2013 (l)	12.28	12.51	2.89	168
GDP (current US$) 2011 (l)	24.55	24.27	2.12	173
GDP (current US$) 2013 (l)	24.66	24.57	2.11	169
Secondary enrollment, percent 2011	80.10	88.85	27.41	127
Tertiary enrollment, percent 2011	40.27	36.55	28.61	123
Population 2011 (l)	15.95	16.03	1.70	179
Population 2013 (l)	15.99	16.06	1.70	174
Innovation capacity[a] 2011 (l)	1.14	1.06	0.27	139
Innovation capacity[a] 2013 (l)	1.28	1.25	0.21	142
Expenditure on public education as a percentage of GDP 2010	4.88	4.77	1.88	112
International trade as a percentage of GDP 2008	102.49	89.31	61.61	151

[a] Innovation capacity is measured using a survey administered to World Economic Forum partner institutes. The survey respondents were asked to evaluate the extent to which companies have the capacity to innovate on a scale of 1 (not at all) to 7 (to a great extent). The country scores are averaged from individual responses and weighed according to the country's past performance.

publishing or collaborative coding. We believe that all three datasets are suitable proxies for measuring traditional and digital content creation.

From all per-country records in the dependent and explanatory variables, we filtered out microstates (populations of less than 250,000 inhabitants) since they tend to introduce fluctuating signals with respect to the metrics outlined above and, thus, are detrimental to regression modeling.

Empirical Model and Estimation Methodology

Based on our exploration of the available data and to address our two research questions, this article adopts a twofold research methodology. To investigate the first research question, we measure and visualize content creation in the context of SSA. Using choropleth maps, we analyze the location of content creation through academic articles, collaborative software development activities, and domain registrations. We then compare the SSA metrics to other world regions.

To respond to the second research question, we investigate how the patterns we find relate to the drivers attributed to knowledge economies. Given the findings from our data exploration, we find that an appropriate model to address our second research question is the multiple linear regression model, which uses Ordinary Least Squares as its estimation method.

We monitor multicollinearity by calculating variance inflation factor (VIF) values for each regressor in each model. Where the VIF values rise above desirable levels, we address the problem of multicollinearity by removing the most inflicted of the collinear variables (Wooldridge, 2012). As we are interested in measuring the relative contributions of all explanatory variables, for those models from which a variable is removed due to multicollinearity, we additionally measure the proportional reduction of error (PRE; Judd, McClelland, & Ryan, 2008). The PRE estimates the percentage of variance explained by a predictor of interest by comparing the residual sum of squares of a model with the predictor of interest and a model that omits the predictor of interest. Through this comparison, the PRE allows for an assessment of the relative contribution that each control variable yields on the dependent variable.

As we are interested in comparing the effect size of the different explanatory variables, we compute beta coefficients, which allow us to measure effects in terms of standard deviation units instead of the original units of the regressors (Wooldridge, 2012). The standard OLS framework does not permit judging the relative importance of regressors based on the size of their coefficient. However, comparing the magnitudes of the beta coefficients enables this comparison.

Finally, the residuals of the main models of interest will be mapped to allow a comparison among countries and between the SSA region and other world regions. The model residuals indicate whether a particular country's level of content creation is higher or lower than predicted by the model that takes into account the country's level of attainment in terms of the explanatory variables.

Geographies of Digital Content Creation

To investigate the geographies of content creation across the three dependent variables, we now examine the total number of academic articles, collaborative coding activity, and domain registrations in countries and regions as well as their total numbers standardized by population. Figure 1 displays the distribution of content creation across the world regions. North America and Europe together account for well over half of the content creation in all three dependent variables: 66.4% of academic articles, 78.5% of collaborative coding, and 76.8% of domain registrations derive from those continents. Academic articles are spread slightly more evenly across the continents of Asia, Middle East/North Africa (MENA), and Oceania, while digital content creation is concentrated in developed countries. SSA contributes the smallest share of content to all three categories, providing only 1.1% of academic articles. With 0.5% of collaborative coding and 0.7% of domain registrations, SSA produces an even smaller share of digital content.

Ranking individual countries with respect to the total number of contributions across the three types of content paints a sobering picture of the domination of the high-income countries. The 10 countries with the greatest amount of content production are located in North America, Europe, Asia, and Oceania (see Table 4).

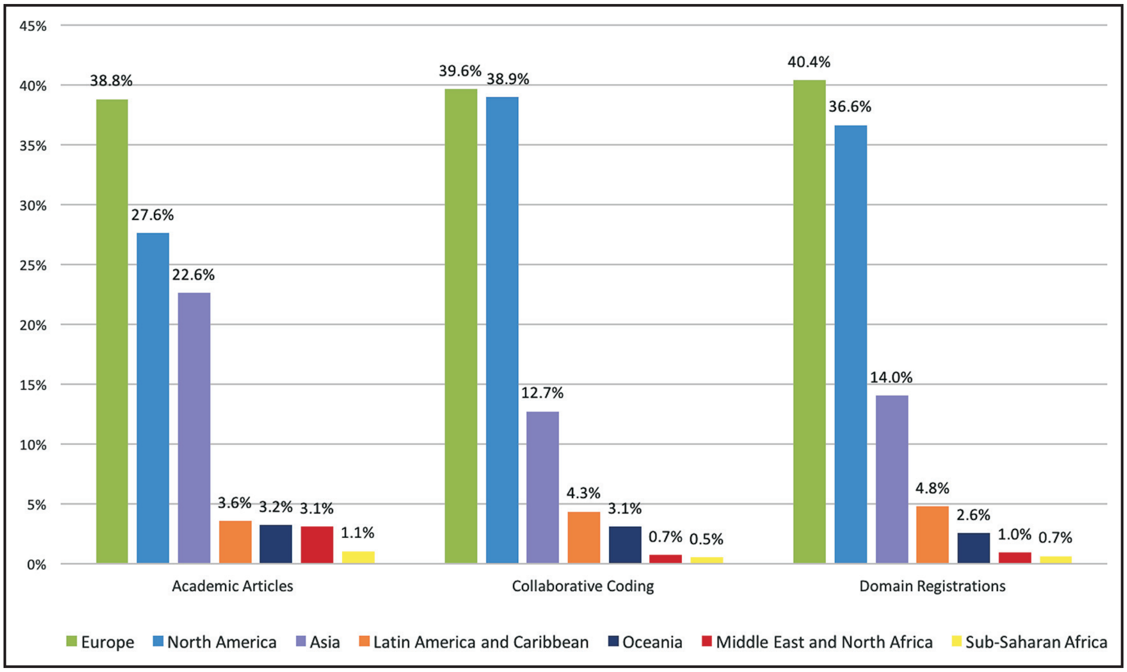

Figure 1. Content creation across continents.

A list of the 50 countries producing the most academic articles includes five countries in Latin America and the Caribbean (LAC), five countries in MENA, and only one country in SSA (South Africa, which was 34th). A similar list pertaining to collaborative coding includes five countries in LAC, one in MENA, and one in SSA (South Africa, 36th). A comparative list of domain registrations includes six countries in LAC, two in MENA, and one in SSA (South Africa, 26th). Of the 50 countries with the fewest academic articles, 23 are in SSA, two are in MENA, and 12 are in LAC. Of the 50 countries with the least activity in collaborative coding, 24 are in SSA, seven are in MENA, and six are in LAC. Finally, among the 50 countries with fewest domain registrations, 31 are in SSA, one is in MENA, and four are in LAC. SSA accounts for approximately half the countries with the least amount of content creation in all categories. Furthermore, although low-income countries provide much less content across all three categories, collaborative coding and domain registrations are even more concentrated in the high-income countries than academic production, and participation from low-income countries and SSA, in particular, is less common.

While comparison across absolute numbers informs us of the total volume of content creation, it is useful to pair that with a standardized measure that informs us of the propensity of content creation across the populations. Ranking individual countries with respect to their per capita content creation reveals that the countries with greatest levels of content contribution are overwhelmingly European. North America and Australia also play a role, but no Asian country reaches the top 10 in any content category (see Table 4).

Considering the most productive countries in terms of their per capita content creation suggests geographies even more clustered in Europe than looking at total numbers. We can look at how the more prominent role of Europe and high-income countries compares with the rest of the world in more detail in the choropleth maps shown in figures 2–4.

Comparing the three maps reveals that several countries in the MENA region reach the fourth quintile of academic articles, while the majority fall within the second and third quintiles of digital content creation. In SSA, the level of individual countries' content production falls within the two lowest quintiles more often in the case of collaborative coding and domain registrations than with academic articles.

Table 4. Country Rankings of Content Creation Across the Three Categories.

Rank	Country	Region	Number	Rank	Country	Region	Number per 1,000 capita
Academic Articles							
1	United States	NOAM	432,557	1	Switzerland	Eur	3.36
2	China	Asia	151,403	2	Denmark	Eur	2.56
3	United Kingdom	Eur	126,192	3	Iceland	Eur	2.53
4	Germany	Eur	104,632	4	Ireland	Eur	2.45
5	Japan	Asia	78,879	5	Sweden	Eur	2.34
6	France	Eur	71,181	6	Netherlands	Eur	2.21
7	Canada	NOAM	64,244	7	Australia	Ocea	2.21
8	Italy	Eur	60,579	8	Norway	Eur	2.16
9	Spain	Eur	53,136	9	United Kingdom	Eur	2.00
10	Australia	Ocea	49,324	10	Finland	Eur	1.97
Collaborative Coding							
1	United States	NOAM	12,169,099	1	Switzerland	Eur	62.39
2	United Kingdom	Eur	2,543,996	2	New Zealand	Ocea	56.58
3	Germany	Eur	2,263,202	3	Sweden	Eur	56.53
4	France	Eur	1,574,058	4	Norway	Eur	53.65
5	China	Asia	1,557,115	5	Iceland	Eur	52.95
6	Canada	NOAM	1,488,541	6	Netherlands	Eur	51.90
7	Japan	Asia	1,154,058	7	Finland	Eur	45.57
8	Russia	Eur	920,458	8	Canada	NOAM	42.34
9	Netherlands	Eur	872,157	9	Denmark	Eur	41.22
10	Australia	Ocea	832,794	10	United Kingdom	Eur	39.68
Domain Registrations							
1	United States	NOAM	82,385,888	1	Netherlands	Eur	475.81
2	Germany	Eur	24,781,840	2	Luxembourg	Eur	347.70
3	United Kingdom	Eur	17,878,584	3	Switzerland	Eur	338.43
4	China	Asia	16,645,066	4	Denmark	Eur	312.35
5	Netherlands	Eur	7,995,636	5	Germany	Eur	307.29
6	France	Eur	7,545,587	6	United Kingdom	Eur	278.89
7	Canada	NOAM	7,140,787	7	United States	NOAM	260.31
8	Russia	Eur	6,582,170	8	Iceland	Eur	257.86
9	Australia	Ocea	5,478,431	9	Australia	Ocea	236.90
10	Japan	Asia	5,081,194	10	Austria	Eur	230.30

Drivers of the Uneven Geographies of Digital Content Creation

To explore our second research question regarding drivers of these uneven distributions, we employ a set of multiple regression models. We then analyze the model residuals, which helps us better understand the uneven geographies of content creation.

We begin the analysis of each dependent variable with the following base specification:

$$\mathbf{Y}_i\,(l) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \text{ Fixed Broadband Connections } (l) + \boldsymbol{\beta}_2 \text{ GDP } (l) + \boldsymbol{\beta}_3 \text{ Enrollment Rate} + \boldsymbol{\beta}_4 \text{ Population } (l) + \mathbf{u}_i$$

The beta coefficients of the base models indicate that GDP has a stronger effect on academic articles than broadband connectivity, but that the two variables seem to have similar effects on collaborative coding and domain registrations. However, in each of the base specifications, the variables measuring GDP and broadband connectivity suffer from high collinearity. As the variable measuring GDP has the highest VIF value, and since we are interested in studying whether broadband connectivity seems to be a significant predictor of changes in

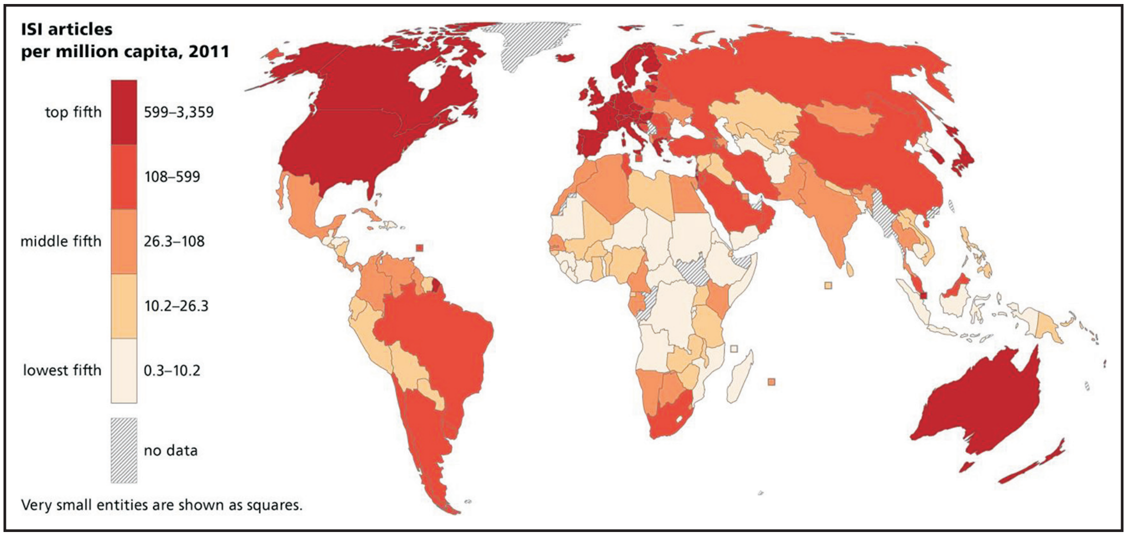

Figure 2. Academic articles per million capita in 2011.

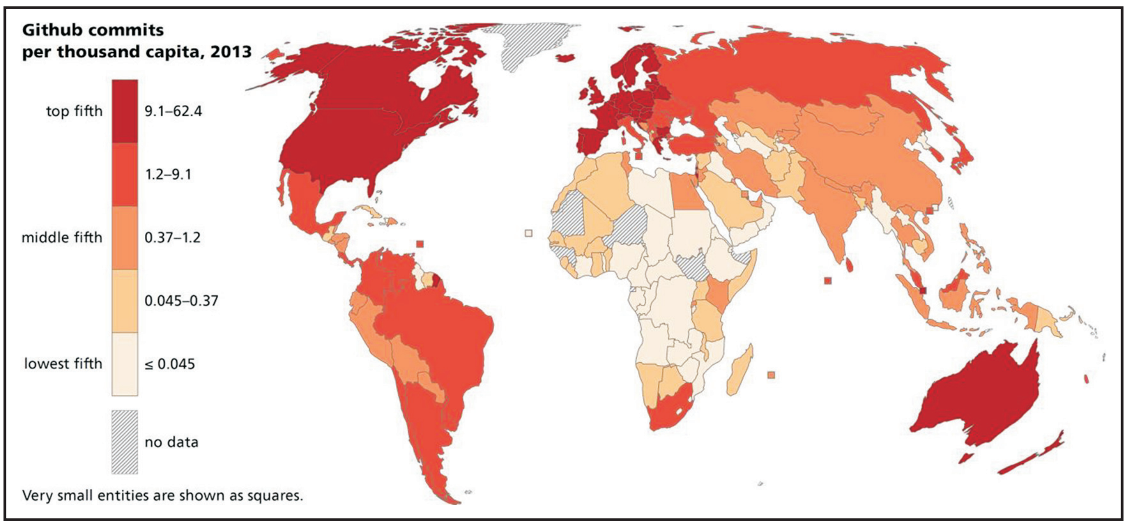

Figure 3. GitHub commits per thousand capita in 2013.

the dependent variables, we remove the GDP variable from subsequent analysis. Given that we include GDP as a regressor only in the base models, and to further investigate the relative effects of GDP and broadband connectivity on the dependent variables, we also analyze the PRE. Table 5 shows the PRE and how the residual sum of squares (RSS) on each of the three dependent variables explained by the base model changes when the variables measuring broadband connectivity and GDP are omitted. We observe that relative to broadband connections, GDP explains a larger part of the variance in academic articles than in the two other content categories. Thus, the PRE analysis confirms the finding from the beta coefficient comparison and suggests that broadband is a more important predictor in changes of collaborative coding and domain registrations than GDP.

Following removal of the variable measuring GDP, we complement the base models by adding a variable that measures innovation capacity as we suspect that innovation might be an important predictor of changes in the dependent variables. We estimate the following specification:

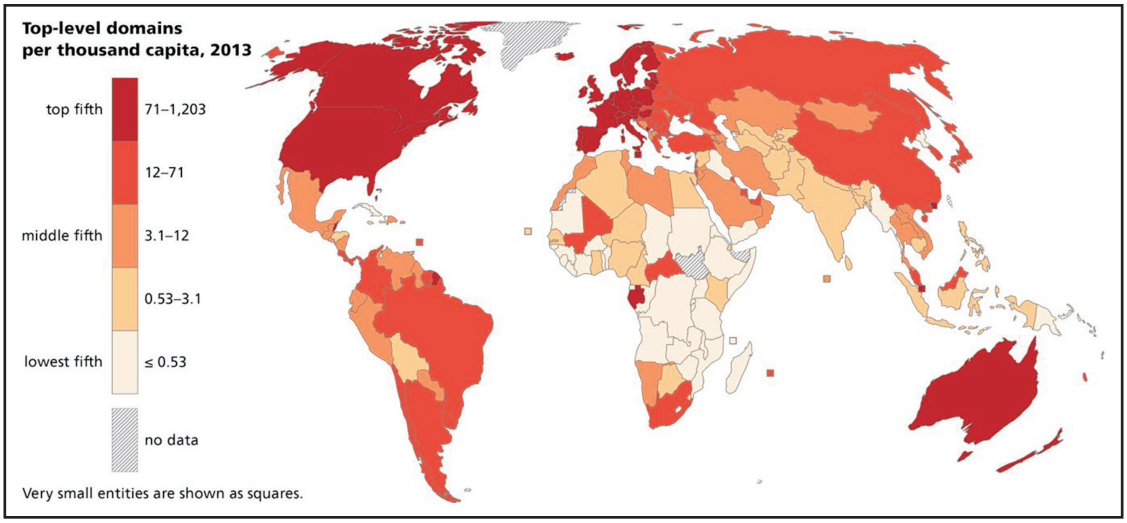

Figure 4. Domain registrations per thousand capita in 2013.

Table 5. Model Comparison Through Proportional Reduction of Error (PRE).

	Model 1 RSS with Both GDP (l) and Broadband Connections (l)	Model 1 RSS with GDP (l) and Excluding Broadband Connections (l)	Model 1 RSS with Broadband Connections (l) and Excluding GDP (l)
Academic articles	64.695 [113]	74.687 (0.134) [115]	85.797 (0.280) [114]
Collaborative coding	306.098 [121]	352.869 (0.133) [123]	333.985 (0.083) [122]
Domain registrations	169.160 [122]	200.600 (0.157) [123]	193.107 (0.121) [123]

Note: PRE values appear in parentheses and the number of observations appear in square brackets. A higher relative PRE value indicates that the omitted variable is a more important predictor.

$$\mathbf{Y}_i\,(l) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \text{ Fixed Broadband Connections (l)} + \boldsymbol{\beta}_2 \text{ Enrolment Rate} + \boldsymbol{\beta}_3 \text{ Population (l)} + \boldsymbol{\beta}_4 \text{ Innovation Capacity (l)} + \mathbf{u}_i$$

Removing the variable that measures GDP and adding innovation capacity stabilizes the model that estimates changes in academic articles as VIF values decrease, but reduces the model fit slightly from 0.918 to 0.893 (see Table 6). The second specification includes tertiary enrollment rate, population, fixed broadband connections, and innovation capacity as significant estimators of changes in the dependent variable. Observing the beta coefficients suggests that population has the strongest relationship to academic articles with an estimator of 0.382, and tertiary enrollment follows with only a slightly smaller effect size, at 0.328. The beta coefficients of fixed broadband connections and innovation capacity have smaller beta coefficients at 0.290 and 0.263, respectively.

Estimating the second specification for collaborative coding decreases VIF values. As a result, the model fit improves slightly from 0.778 to 0.780. The second specification includes fixed broadband connections, population, and innovation capacity as significant estimators of changes in the dependent variable. Observing the beta coefficients suggests that fixed broadband connections has the strongest relationship to collaborative

Table 6. Regression Analysis of the Determinants of the Dependent Variables.

	Academic Articles		**Collaborative Coding**		**Domain Registrations**	
	Model 2	**Model 3**	**Model 2**	**Model 3**	**Model 2**	**Model 3**
Main Explanatory Variables						
Fixed broadband connections (l)	0.277*** (0.077) [0.290] VIF: 4.86	0.254*** (0.095) [0.258] VIF: 5.42	0.754*** (0.211) [0.588] VIF: 7.64	0.759*** (0.226) [0.596] VIF: 7.08	0.805*** (0.127) [0.702] VIF: 7.59	0.766*** (0.129) [0.666] VIF: 7.78
Tertiary enrollment	0.029*** (0.005) [0.328] VIF: 3.04	0.034*** (0.007) [0.372] VIF: 4.10				
Secondary enrollment			0.011 (0.012) [0.092] VIF: 5.08	0.007 (0.013) [0.063] VIF: 5.19	0.013 (0.009) [0.119] VIF: 4.78	0.012 (0.008) [0.113] VIF: 5.02
Population (l)	0.553*** (0.085) [0.382] VIF: 2.36	0.617*** (0.105) [0.397] VIF: 2.45	0.395** (0.194) [0.209] VIF: 4.81	0.502*** (0.187) [0.259] VIF: 4.13	0.154 (0.148) [0.090] VIF: 4.56	0.336** (0.138) [0.191] VIF: 5.05
Innovation capacity (l)	2.321*** (0.376) [0.263] VIF: 2.15	2.117*** (0.483) [0.240] VIF: 2.75	2.976*** (0.795) [0.199] VIF: 1.52	1.837* (1.023) [0.123] VIF: 1.71	2.441*** (0.528) [0.184] VIF: 1.64	1.915** (0.769) [0.146] VIF: 1.85
Additional Control Variables						
Public spending on education as a percentage of GDP		0.031 (0.068) [0.019] VIF: 1.40		0.391*** (0.122) [0.192] VIF: 1.24		0. (0.097) [0.113] VIF: 1.26
International trade as a percentage of GDP						0.002 (0.002) [0.033] VIF: 1.36
Mean VIF Score						
Mean VIF	3.10	3.22	4.76	3.87	4.64	3.72
Measures of Fit						
Constant	−9.261*** (0.995)	−10.147*** (1.262)	−11.619*** (2.060)	−13.265*** (2.099)	−5.750*** (1.721)	−8.556*** (1.645)
R2	0.893	0.902	0.780	0.808	0.886	0.911
Number of Observations	100	76	108	81	107	76

Notes: Robust standard errors appear in parentheses; beta coefficients appear in square brackets.
** Significant at 10%; ** Significant at 5%; *** Significant at 1%.*

coding, with a beta coefficient of 0.588; population and innovation capacity follow with a smaller effect size at 0.209 and 0.199, respectively.

In the case of domain registrations, the second model is similarly more stable as VIF values decrease, and the model fit improves from 0.839 to 0.886. The second specification includes fixed broadband connections and innovation capacity as significant estimators of changes in the dependent variable. The beta coefficients suggest that fixed broadband connections has the strongest relationship to domain registrations with an estimator of 0.702, and innovation capacity follows with a smaller effect size at 0.184.

We check the stabilized second specification for robustness by adding a further control variable that measures public spending on education. Additionally, for the model measuring changes in domain registrations, we include a variable that measures international trade to control for countries' differing trade volumes. Thus, we estimate the following third specification:[6]

$$\mathbf{Y}_i\text{ (l)} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1\text{ Fixed Broadband Connections (l)} + \boldsymbol{\beta}_2\text{ Enrolment Rate} + \boldsymbol{\beta}_3\text{ Population (l)}+$$
$$\boldsymbol{\beta}_4\text{ Innovation Capacity (l)} +\boldsymbol{\beta}_5\text{ Public Expenditure on Education as a Percentage of GDP} +$$
$$\boldsymbol{\beta}_6\text{ International Trade as a Percentage of GDP}^6 + \mathbf{u}_i$$

Adding the new control variable to the analysis of academic articles improves the model fit from 0.893 in Model 2 to 0.902 in Model 3. Model stability is good at mean VIF 3.22. The new control variable does not enter the regression as a significant predictor of changes in the dependent variable. Further, adding this new control variable does not significantly change the size of the estimators in the base model, which suggests a robust base model.

Similarly, for collaborative coding, adding the new control variable improves the model fit from 0.780 in Model 2 to 0.808 in Model 3. The third model suggests that public spending on education is a significant estimator of changes in the number of GitHub commits. Adding this new control variable does not significantly change the size of the estimators in the base model, suggesting a robust base model. However, adding the new control variable renders the variable that measures innovation capacity significant only at the 10% level.

Finally, for domain registrations, adding the two new control variables improves the model fit from 0.886 in Model 2 to 0.911 in Model 3. The third model suggests that public spending on education and population are significant predictors of changes in the dependent variable. However, adding the new control variables does not significantly change the size of the estimators in the base model, which suggests a robust base model.

However, results from the third specification must be examined with caution across all dependent variables because the number of observations is lower in each regression. Owing to this, and given that the third specification suggests a robust base model for all three regressions, in the following section we map the regression results for the second specification for each dependent variable.

The regression models suggest that GDP, broadband connectivity, and innovation capacity are key predictors of all three types of content creation. However, these results imply that the factors often framed as catalysts in the transformation into, and development of, a knowledge economy do not relate to academic articles, collaborative coding activity, and domain registrations uniformly. The regression coefficients in Model 1 show that while a percent increase in GDP is related to approximately a 0.6% increase in all three dependent variables, a percent increase in connectivity is related to a 0.11% increase in academic articles, but nearly fivefold increases in collaborative coding and domain registrations (0.51% and 0.46%, respectively). Further, the regression coefficients in Model 2 indicate that a percent increase in broadband connectivity is related to a 0.28% increase in academic articles, while for collaborative coding and domain registrations, the associated change is nearly twice or three times as much (0.75% and 0.84%, respectively). On the other hand, a percent increase in innovation capacity is related to larger changes at 2.32%, 2.98%, and 2.41% increases for academic articles, coding, and domain registrations, respectively. In this vein, merely increasing connectivity without supporting GDP growth or increasing innovation capacity might not allow countries to leapfrog to higher levels of digital content creation and has an even smaller effect on production of academic articles.

Mapping Regression Residuals

The choropleth maps in figures 5–7 show the residuals from the second model specifications and investigate whether regions and specific countries seem to under- or overperform relative to the model expectations.

The residual maps of the models suggest the presence of a divide between high-income and low-income

6. The variable that controls for International Trade as a Percentage of GDP is only included for the model that measures changes in domain registrations.

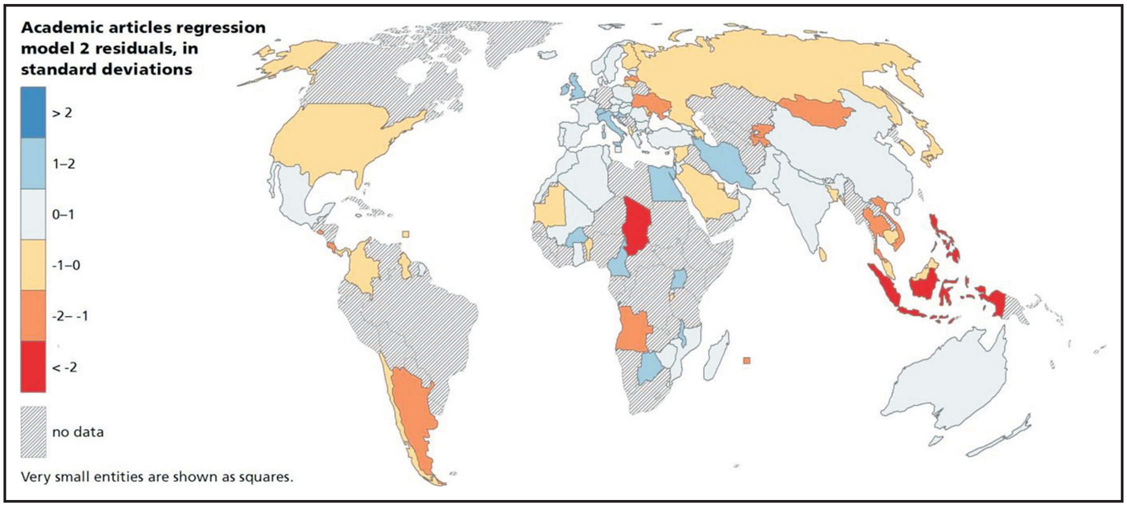

Figure 5. Residual map of academic articles.

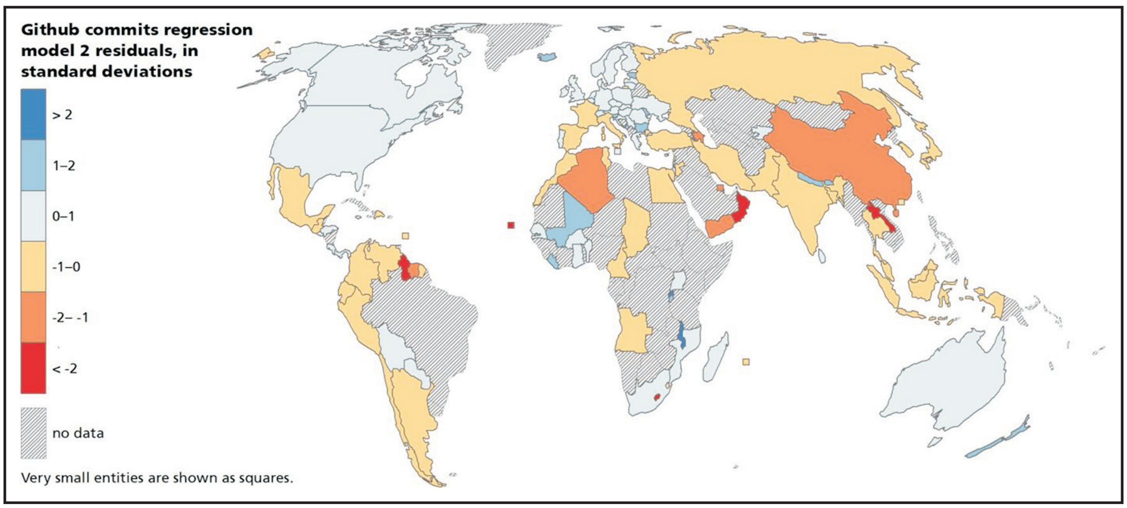

Figure 6. Residual map of GitHub commits.

countries, where the former tend to produce more content and the latter less content than would be predicted by the model, taking into account the country's level of attainment in terms of broadband connectivity, education, innovation capacity, and public spending on education. This difference appears starker for the two digital content categories. Various SSA countries produce more academic articles than the model suggests, but only a handful of them produce more digital content. This pattern further highlights deepening divides in the digital production of content and suggests that the presence of enabling drivers may not automatically translate to digital content creation in the region.

Conclusions

Changes in the ways that information and knowledge are created, used, and reproduced, and how they are enabled by new information technologies, are driving developments in economic and social structures and interactions worldwide. Recent advances in connectivity have sparked hope for the democratization of information and knowledge production, particularly in SSA in the wake of rapidly growing connectivity and

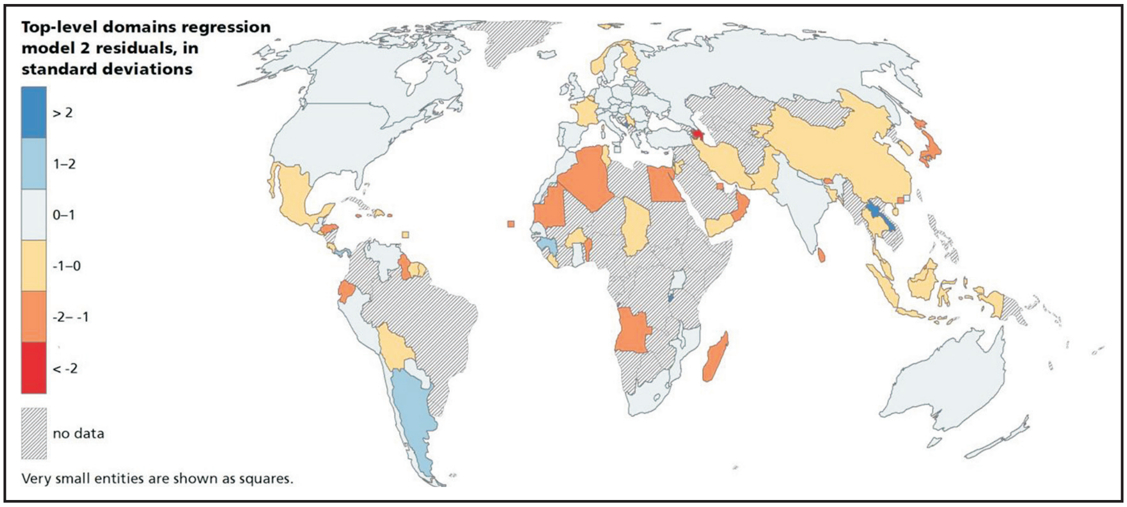

Figure 7. Residual map of top-level domains.

telecom industry development. However, empirical evidence on digital content creation is only beginning to emerge, and there is scant research about the extent of these developments in SSA.

To address this research gap, this article has investigated the geographies of knowledge-intensive content production within SSA countries and between SSA and other world regions. Further, it has studied whether the drivers commonly associated with the knowledge economy explain these geographies. We find that the geographies of collaborative coding and domain registrations are more uneven than the spatial distribution of academic authoring. This runs contrary to the expectation of contemporary digitally mediated content being more evenly geographically distributed than traditional content.

Our results suggest that the factors often framed as catalysts in the transformation into, and development of, a knowledge economy do not relate uniformly to academic articles, collaborative coding activity, and domain registrations. While connectivity plays a role in all three categories, it seems to have a strong effect only on digital content creation. Conversely, the production of academic articles is more strongly related to GDP than to connectivity. Innovation capacity appears to have a positive relationship to all three content types. Education as a topically narrower variable appears, perhaps unexpectedly, to be related only to variance in academic articles.

These observations suggest an important conclusion for policy makers in SSA: While connectivity is clearly an important enabler of digitally mediated content creation, merely increasing connectivity might not allow SSA countries to leapfrog to higher levels of digital content creation. Wealth, innovation capacity, and public spending on education matter as well. While the growth in telecommunications might be slackening the continent's reliance on extractive industries and agriculture, transformation into a knowledge economy requires far more concentrated effort than simply increasing Internet connectivity.

The role of geography continues to matter in the era of digitally mediated content production. Collaborative coding and domain registrations remain characterized by sticky geographies (Markusen, 1996), interestingly, more so than academic articles, which seemed initially to require a more geographically clustered and institutionally heavy set of geographically sticky factors. To the extent that digital information will continue to gain a more prominent role in contemporary knowledge economies, the uneven geographies of digitally mediated content indicate a worryingly diminishing role for SSA vis-à-vis other world regions.

Given the central role of ICTs and digitally mediated activities for the knowledge economy, it is surprising that digital participation and content creation are rarely included in related studies. As patterns of information- and knowledge-rich activities are notoriously difficult to measure, we hope that information about digital participation and content creation offers a complementary data source for studies in this field. ■

Acknowledgments

We would like to thank Professor Juan Pablo Alperin of Simon Fraser University for sharing his data set on academic articles.

Funding

The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013) / ERC Grant Agreement n. 335716.

Sanna Ojanperä, Researcher, Oxford Internet Institute, University of Oxford, United Kingdom. sanna.ojanpera@oii.ox.ac.uk

Mark Graham, Professor of Internet Geography, Oxford Internet Institute, University of Oxford and Faculty Fellow, The Alan Turing Institute, United Kingdom. mark.graham@oii.ox.ac.uk

Ralph K. Straumann, Visiting Researcher, Oxford Internet Institute, University of Oxford, United Kingdom. ralph.straumann@oii.ox.ac.uk

Stefano De Sabbata, Lecturer in Quantitative Geography, Department of Geography, University of Leicester, United Kingdom. s.desabbata@leicester.ac.uk

Matthew Zook, Professor of Economic and Information Geography, Department of Geography, University of Kentucky, United States. zook@uky.edu

References

Adams, P., Capiluppi, A., & Boldyreff, C. (2009, September). *Coordination and productivity issues in free software: The role of Brooks' Law.* Paper presented at IEEE International Conference on Software Maintenance, Edmonton, AB. doi:10.1109/ICSM.2009.5306308

Afele, J. S. (2002). *Digital bridges: Developing countries in the knowledge economy.* Hershey, PA: Idea Group Inc.

Arora, A., & Gambardella, A. (2004). *The globalization of the software industry: Perspectives and opportunities for developed and developing countries* (NBER Working Paper No. 10538). Cambridge, MA: National Bureau of Economic Research.

Balland, P., & Rigby, D. (2015). *The geography and evolution of complex knowledge* (Papers in Evolutionary Economic Geography No. 1502). Utrecht, the Netherlands: Utrecht University.

Blankley, W. O., & Booyens, I. (2010). Building a knowledge economy in South Africa. *South African Journal of Science, 106*(11–12), 1–6.

Britz, J. J., Lor, P. J., Coetzee, I. E. M., & Bester, B. C. (2006). Africa as a knowledge society. *International Information & Library Review, 38*(1), 25–40.

Carmody, P. (2013). A knowledge economy or an information society in Africa? Thintegration and the mobile phone revolution. *Information Technology for Development, 19*(1), 24–39.

Casado-Lumbreras, C., Colomo-Palacios, R., Ogwueleka, F., & Misra, S. (2014). Software development outsourcing: Challenges and opportunities in Nigeria. *Journal of Global Information Technology Management, 17*(4), 267–282.

Castells, M. (2003). *The power of identity* (2nd ed.). Oxford, UK: Blackwell.

Cooke, P. (2001). Regional innovation systems, clusters, and the knowledge economy. *Industrial and Corporate Change, 10*(4), 945–974.

Friederici, N., Ojanperä, S., & Graham, M. (2017). The impact of connectivity in Africa: Grand visions and the mirage of inclusive digital development. *Electronic Journal of Information Systems in Developing Countries*, *79*(2), 1–20.

Gaibi, Z., Maske, A., & Moraje, S. (2010, June). Africa's path to growth: Sector by sector. *McKinsey Quarterly Economic Studies.* Retrieved from http://www.mckinsey.com/global-themes/middle-east-and-africa/africas-path-to-growth-sector-by-sector

GitHub. (2015). *Build software better, together.* Retrieved from https://github.com/about

Graham, M., De Sabbata, S., & Zook, M. A. (2015). Towards a study of information geographies: (im)mutable augmentations and a mapping of the geographies of information. *Geo: Geography and Environment, 2*(1), 88–105.

Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014). Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Annals of the Association of American Geographers, 104*(4), 746–764.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature, 28*(4), 1661–1707.

Jaffe, A. B., & Trajtenberg, M. (2002). *Patents, citations, and innovations: A window on the knowledge economy.* London, UK: MIT Press.

Judd, C., McClelland, G., & Ryan, C. (2008). *Data analysis: A model comparison approach* (2nd ed.). New York, NY: Routledge.

Juma, C., & Agwara, H. (2006). Africa in the global knowledge economy: Strategic options. *International Journal of Technology and Globalisation, 2*(3–4), 218–231.

Kalliamvakou, E., Singer, L., Gousios, G., German, D., Blincoe, K., & Damian, D. (2014, May 31–June 1). *The promises and perils of mining GitHub.* Paper presented at the 11th Working Conference on Mining Software Repositories, Hyderabad, India. doi:10.1145/2597073.2597074

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences* [electronic resource]. Los Angeles, CA: SAGE Publications.

Kolo, J. (2009). The knowledge economy: Concept, global trends and strategic challenges for Africa in the quest for sustainable development. *International Journal of Technology Management, 45*(1–2), 27–49.

Lacity, M. C., Willcocks, L. P., & Rottman, J. W. (2008). Global outsourcing of back office services: Lessons, trends, and enduring challenges. *Strategic Outsourcing: An International Journal, 1*(1), 13–34.

Lauf, E. (2005). National diversity of major international journals in the field of communication. *Journal of Communication, 55*(1), 139–151.

Lesotho Ministry of Communications, Science and Technology. (2005, March 4). *ICT policy for Lesotho*. Retrieved from http://www.ist-africa.org/home/files/Lesotho_ICTPolicy_2005.pdf

Lima, A., Rossi, L., & Musolesi, M. (2014). *Coding together at scale: GitHub as a collaborative social network*. Paper presented at 8th AAAI International Conference on Weblogs and Social Media, Ann Arbor, MI. Retrieved from arXiv:1407.2535

Lundvall, B., & Johnson, B. (1994). The learning economy. *Journal of Industry Studies, 1*(2), 23–42.

Marker, P., McNamara, K., & Wallace, L. (2002). *The significance of information and communication technologies for reducing poverty.* London, UK: Department for International Development.

Markusen, A. (1996). Sticky places in slippery space: A typology of industrial districts. *Economic Geography, 72*(3), 293–313.

McNamara, K. (2003). *Information and communication technologies, poverty and development: Learning from experience.* Washington, DC: The World Bank.

Murenzi, R., & Hughes, M. (2006). Building a prosperous global knowledge economy in Africa: Rwanda as a case study. *International Journal of Technology and Globalisation, 2*(3–4), 252–267.

Nyamnjoh, F. (2004). From publish or perish to publish and perish: What Africa's 100 best books tell us about publishing Africa. *Journal of Asian and African Studies, 39*(5), 331–355.

O'Brien, R. (1992). *Global financial integration: The end of geography.* London, UK: Council on Foreign Relations Press.

Ondari-Okemwa, E. (2007, October 1). Scholarly publishing in sub-Saharan Africa in the twenty-first century: Challenges and opportunities. *First Monday, 12*(10). Retrieved from http://firstmonday.org/article/view/1966/1842

Powell, M. (2001). Knowledge, culture & the Internet in Africa: A challenge for political economists. *Review of African Political Economy, 28*(88), 241–260.

Powell, W., & Snellman, K. (2004). The knowledge economy. *Annual Review of Sociology, 30*(1), 199–220.

Republic of Ghana. (2003, June). *The Ghana ICT for accelerated development (ICT4AD) policy.* Retrieved from https://cdn.modernghana.com/images/content/report_content/ICTAD.pdf

Sonn, J., & Storper, M. (2008). The increasing importance of geographical proximity in knowledge production: An analysis of US patent citations, 1975–1997. *Environment and Planning A, 40*(5), 1020–1039.

Sörlin, S., & Vessuri, H. (2006). *Knowledge society vs. knowledge economy.* Basingstoke, UK: Palgrave Macmillan.

Takhteyev, Y., & Hilts, A. (2010). *Investigating the geography of open source software through Github.* Manuscript submitted for publication.

Udo, G. J., & Edoho, F. M. (2000). Information technology transfer to African nations: An economic development mandate. *The Journal of Technology Transfer, 25*(3), 329–342.

Waghid, Y. (2002). Knowledge production and higher education transformation in South Africa: Towards reflexivity in university teaching, research and community service. *Higher Education, 43*(4), 457–488.

Wooldridge, J. (2012). *Introductory econometrics: A modern approach* (5th ed.). Delhi, India: South-Western College Publishing.

World Bank. (2012). *Knowledge economy index* [Data file]. Retrieved from https://knoema.com/WBKEI2013/knowledge-economy-index-world-bank-2012

Zegeye, A., & Vambe, M. (2006). Knowledge production and publishing in Africa. *Development Southern Africa, 23*(3), 333–349.

Zook, M. A. (2001). Old hierarchies or new networks of centrality? The global geography of the Internet content market. *American Behavioral Scientist, 44*(10), 1679–1696.