Measuring the Hidden Contours of The Global Knowledge Economy with a Digital Index

Sanna Ojanperä, Oxford Internet Institute, University of Oxford. sanna.ojanpera@oii.ox.ac.uk
Mark Graham, Oxford Internet Institute, University of Oxford. mark.graham@oii.ox.ac.uk
Matthew Zook, University of Kentucky. zook@uky.edu

Short abstract:

The *Digital Knowledge Economy Index*, combines traditional data sources with bespoke data on capacities and skills (measured via content-creation and participation on digital platforms) to provide a revealing view of where developing countries (with a particular focus on Sub-Saharan Africa) fit into the world's digital knowledge economy.

Long abstract:

Taking advantage of the 'information revolution' is a priority in various national development strategies. Eager to tap into economic and social opportunities potentially afforded by increased access to digital information, many governments of developing countries have envisioned policies that guide their transformation into so-called 'knowledge economies'. However, despite its importance to contemporary economic development, the concept itself is rarely clearly defined, operationalised, or effectively measured. The few indices that have measured the state of knowledge and information economies around the world employ significantly different sets of variables from one another, further adding to the vagueness of the concept. These indices rely on the accuracy and cross-sectional as well as longitudinal representativeness of their data sources, which can be called into question in low-income contexts. We thus propose the construction of a *Digital Knowledge Economy Index*, which will additionally account for capacities and skills that are quantifiable via measuring content-creation and participation directly through digital platforms, such as the code-sharing platform GitHub, encyclopaedia Wikipedia and domain name registrations. With this approach, conventional data sources - national statistics and expert surveys - can be complemented by data that is collected via the internet and that reflects the underlying digital content creation, capacities and skills of the population. An index that combines traditional and novel data sources may provide a more revealing view of the status of the world's digital knowledge economy and highlight where the availability of digital resources may actually reinforce inequalities in the age of data.

Keywords: digital divide, information geographies, geographies of knowledge, international development, measurement, evaluation, GitHub, Wikipedia, domain registrations

1. Introduction

The recent increase in the availability and capacity of information and communication technologies combined with a shift in the global economic restructuring towards a focus on services, technology and human capital has generated excitement around the 'Information Revolution'. The vast expansion of digitally recorded data as well as a simultaneous extension of both processing power and storage capacity have inspired many national development strategies to take advantage of these developments through policy action. Eager to tap into economic and social opportunities potentially afforded by increased access to digital information, many governments of developing countries have envisioned policies that guide their transformation

into so-called 'knowledge economies'. While the concept of a knowledge economy and its relation to the associated ideas of 'information society', 'information economy', 'knowledge society', and 'digital economy' do not have clear universally accepted definitions, a greater reliance on human capital over natural resources or material factors and a focus on information technology are amongst the central tenets of the term.

In the literature, various high-income countries are either considered to be already-existing knowledge economies (Bałtowski & Pastuszak, 2008; Huggins, 2008; Nguyen & Pham, 2011) or to be part of the global knowledge economy. Conversely, the majority of low-income (/developing) countries are considered as places that are seeking to transform into ones (Carmody, 2013; Kolo, 2009; Piaggesi & Chea, 2011). Many countries in Sub-Saharan Africa, in particular, seem eager to tap into the new opportunities afforded by the information revolution, and both ICT sector policies as well as voices in the donor and private sectors are optimistic about the potentials of such a transformation (Friederici, Ojanperä, & Graham, 2016). For instance, the Leshotho policy states: "ICTs can have a wizard effect on speeding the development process of a country" (Friederici et al., 2016). A common expectation (which tends to fuse developmental teleologies of modernization and technological determinism) sees Sub-Saharan African countries leapfrogging past less efficient technologies and industries and straight into the more advanced ones with a focus on human capital (Graham et. al. 2015).[i] Beyond national and regional case studies, a few indices (created by institutions such as the OECD and World Bank) seek to comparatively measure how close different countries are to having a knowledge economy. Where data from Sub-Saharan African countries is included in these indices, the region consistently ranks lowest.

The indices that measure the state of knowledge (and information) economies around the world employ significantly different sets of variables from one another, as well as differing methodologies: further adding to the vagueness of the concept. A major part of a country's score or rank on the indices is made up by data collected by national statistical institutions. These indices rely on the accuracy and cross-sectional as well as longitudinal representativeness of their data sources, which may be called into question in low-income contexts and Sub-Saharan Africa in particular, where national statistical institutions are often weak. Previous research indicates that the region suffers from a 'statistical tragedy', whereby weak capacity, inadequate funding, and lack of coordination have resulted in unreliable statistical estimates (Devarajan, 2013; Jerven, 2013a, 2013b; Lehohla, 2008)). Further, many of the variables comprising the existing indices are collected using surveys that generalize the finding from a subsample to the larger population, while very few of them feature variables that directly measure total volume of knowledge-intensive activities in the country.

Given the importance that academic, public, donor and private actors place on the idea of Sub-Saharan Africa's transformation into knowledge economies and the substantial diversity within the indices measuring the attainment of this transformation, we propose a reflexive enquiry into the ways that we measure knowledge economies. While many of the existing indices do state their general limitations related to thematical scope and geographical and temporal coverage, they are often created to work as easy-to-use interactive tools for quick identification of challenges and opportunities related to a transition to a knowledge-based economy. The applied natures of the indices lend themselves to quick benchmarking, but discourage deeper reflection about the particular measurement choices that were made in constructing them. The description of the indicators and their scores tend to be accompanied with only narrow discussion about the limitations inherent in the choice of variables or the measurement of those variables. Further,

the index methodology and normalization procedures in particular are often presented matter-of-factly.

We propose evaluating these indices from the point of view that the measurement choices carried out in constructing them are not neutral or value-free purely technical considerations, but conscious choices that result in the index scoring distributed a certain way between the countries included in it. Deeper discussion about the value-laden nature of these indices is currently missing, and we consider that they should be accompanied with it, as their quick and easy interpretations do influence perceptions, discourse and decision-making as well as funding decisions in the countries that aspire to transform into knowledge economies. For instance, Chen and Dahlman, the creators of the World Bank Knowledge Economy Index, report that results from the index have been "routinely used to initiate policy dialogue within a country and to identify issues for further investigation" (2006, pp. 14–15). We pay particular attention into studying the position of Sub-Saharan African countries in each index, as given the region's interest in transition to a knowledge economy, academia, policymakers, donors, and businesses are likely to take note of the lessons that these indices offer them.

While the existing indices tend to measure a combination of attainment in education, innovation, economic competitiveness, and infrastructure, they very rarely include variables estimating knowledge-intensive user activity or participation. The global proliferation of computing devices and digital connectivity have enabled many activities to be carried out over the internet, and leave behind a wealth of digital traces about the users and the tasks they perform. Given the centrality of technology and human capital in the knowledge economy, we maintain that a measurement of knowledge economies should also feature an estimation of actual knowledge-rich digital user activity. Accounting for the measurable activity on digital platforms offers a valuable proxy of the attainment of digital skills and knowledge, which would be difficult to measure otherwise. Comparing these attributes between countries makes visible the actual level of digital user activity and content creation and provides the conceptualization of knowledge economy with a new pool of evidence. With this approach, conventional data sources (i.e. national statistics and expert surveys) can be complemented by data that is collected over the internet and better reflects the underlying digital content creation, capacities and skills of the population. An index that combines traditional and novel data sources may provide a more revealing view of the status of the world's digital knowledge economy and highlight where the availability of digital resources may actually reinforce inequalities.

2. Literature review

The knowledge economy is not only challenging to define conceptually, it is also notoriously difficult to measure or compare between countries. Since the 1990s various assessment frameworks have defined and redefined the dimensions of the knowledge economy, but indices measuring the level of attainment or readiness to compete in the knowledge economy have emerged only over the past decade (see Table 1).

Table 1. Knowledge Economy Frameworks, Indices, and Related Indices Focusing on Digitalization.

Title	Publisher	Date	Notes
Knowledge Economy Frameworks (excluding national frameworks)			
The Knowledge-Based Economy	OECD	1996	

OECD Science, Technology and Industry Scoreboard	OECD	1999, 2001, 2003, 2005, 2007, 2009, 2011, 2013, 2015	
Towards Knowledge-Based Economies in APEC	APEC	2000	
Knowledge Assessment Methodology	World Bank	2006	
Knowledge Economy Indices			
Knowledge Economy Index (KEI)	World Bank	2010, 2012	
Knowledge Index (KI)	World Bank	2010, 2012	
DESI: Digital Economy and Society Index	European Commission	2015, 2016	Preceded by Digital Competitiveness Report published annually from 2009 to 2014
State New Economy Index	Information Technology and Innovation Foundation	2002, 2007, 2008, 2010, 2012, 2014	
Related Indices Focusing on Digitalization			
Digital Evolution Index	The Fletcher School Institute for Business in the Global Context	2008-2013	The index estimates the trajectories over six years rather than annual scores
Industry Digitization Index	Strategy&	2012	
Mapping the European ICT Poles of Excellence: The Atlas of ICT Activity in Europe	European Commission	2014	
Web Index	World Wide Web Foundation	2013 2014	

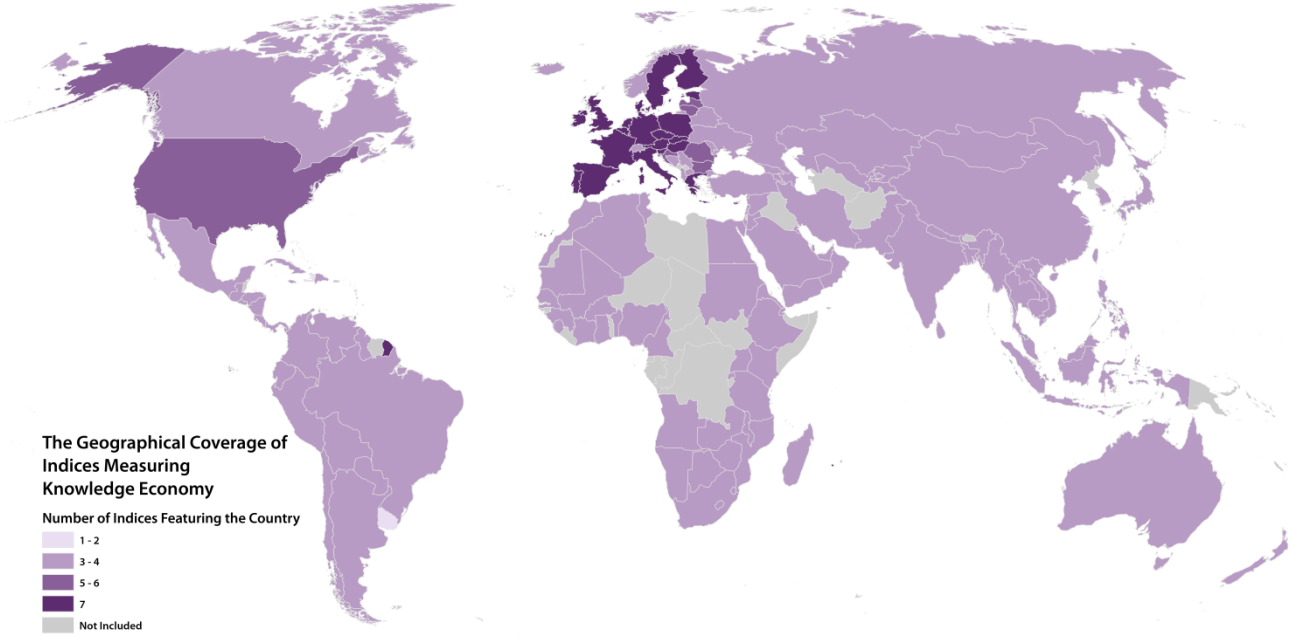
Note: Related groups of indices focusing on entrepreneurship, innovation, and competitiveness, respectively, have not been included in this paper in the interest of brevity, but may be obtained by contacting the authors.

While a few of the indices include countries from Sub-Saharan Africa, only the KI and KEI from the World Bank and the Web Index have a truly global coverage both including 146 countries with 31 from Sub-Saharan Africa. The KI measures a country's attainment of knowledge economy with a score between 10 (best) and 0 (worst) and the KEI uses the same scale to measure a country's ability to compete in the knowledge economy. Amongst African countries, Mauritius ranks highest in the KEI at 67th and with an index score of 5.52, while South Africa tops the KI with a rank of 69 and an index score of 5.11. The average index score between the included Sub-Saharan African countries is 2.18 for KEI and 2.08 for KE in 2012. In comparison, the US ranks 12th in the KEI with a score of 8.77 and 9th in the KI with a score of 8.89.[ii]

The Web Index measures the World Wide Web's contribution to social, economic and political progress in countries across the world on a scale of 0 (worst) to 100 (best). It includes 86 countries with 21 countries from Sub-Saharan Africa. Mauritius ranks highest as 40th with a score of 49.6. The average score for the Sub-Saharan African countries is 22.64. As a comparison, the United States ranks 6th with a score of 94.52. The Sub-Saharan African average score thus measures 25 percent of the US score for the KEI and 24 percent for the Web Index. Among the (only!) three Sub-Saharan African countries included in the Digital Evolution index measuring the evolution of digital economy, South Africa ranks 33rd, while Kenya and Nigeria rank last among the 50 countries included in the study. The index scores vary between 56.21 (best) and 13.77 (worst) and with a score of 30.06, South Africa measures twice the index scores 16.98 for Kenya and 13.77 for Nigeria, while all three fall far behind the US ranking 6th with a score of 51.79. The average index score of these three Sub-Saharan African countries measures 40 percent of the US index score.

In sum, Sub-Saharan African countries tend to sit at the very bottom of the indices. Notable exceptions to this general pattern are Mauritius, South Africa, and to a certain extent also Botswana and Namibia. More worrying is that the majority of the indices measuring knowledge economies do not even include most countries from the region (see Figure 1.). To some extent this could result from paucity of available data, or the questionable reliability of that data. The methodological note accompanying the Knowledge Assement Methodology of the World Bank mentions that countries were not included if data for more than one variable from each of the four subindices was not available (The World Bank, 2013). Some indexes only focused on a particular region of interest (often Europe), and, as such, did not include countries from other regions.

Figure 1. The Geographical Coverage of Indices Measuring Knowledge Economy.



Yet, the limited inclusion of Sub-Saharan Africa in these indices may leave the region having to evaluate domestic opportunities and challenges in light of knowledge economy indicators borrowed from different contexts. Indeed, this has been documented in a recent paper studying African ICT policies which found that despite drastically differing contexts, various policies refer to the experience of high-income Western countries, as predictive of similar developments in Sub-Saharan Africa (Friederici et al., 2016). Given the considerable interest in the region towards transformations into knowledge economies, there is thus a clear need for a knowledge economy index that better covers Sub-Saharan Africa, and includes variables that provide suitable data for estimating the attainment or preparedness for knowledge economy in those places. Designing appropriate knowledge economy policies and projects for the region will ultimately be problematic if the region bases its goals and objectives on the experiences of Western countries.

Existing indices measuring the knowledge economy differ in terms of their variables and methodologies, but all of them have a strong focus on innovation, ICTs, and education. While the groups of related indices focusing on digitalization, entrepreneurship, innovation, and competitiveness navigate towards their particular foci in terms of variable choices, they also largely overlap with the indices fully focused on measuring knowledge economy.

The existing indices exclusively focus on traditional data sources, which often derive from quantitive national aggregates and qualitative expert surveys. As mentioned earlier, these data are often collected by reputable institutions working with national statistical institutions, which in Sub-Saharan Africa are frequently underfunded and may lack capacity in statistical development (Devarajan, 2013; Jerven, 2013a; Lehohla, 2008). Large parts of the region's economic activity operate in the 'informal economy', which has been challenging for the statistical offices to account, suggests Jerven (2013a). After the economic collapse of the 1980s and 1990s or the so called 'lost decades', the continent has struggled to recover from the period of structural adjustment. Further, the need to collect appropriate data for poverty monitoring and for results based management driven by the development community have left national statistical institutes with changing demands and agendas, but lacking implementation strategies in a context of already weak capacities. While many dimensions of the knowledge economy are notably difficult to measure, where consistent quantitative data is not available, existing indices resort to qualitative data derived from expert opinion surveys. The construction of the indices using traditional quantitative and qualitative data results in a measurement of the knowledge economy that for one part derives from quantitative performance data (sometimes of questionable quality) and for another is based on the estimation of experts.

While these indices have been carefully constructed and emphasize the importance of ICTs, hardly any of them utilize variables deriving from direct digital user activity. Such data is used in merely two instances: the Web Index's measurements of dominant social network monthly active users as part of its subindex on 'Relevant Content and Use', and The Digital Evolution Index's measures use of informational websites and social media usage as part of its subindex on 'Internet and Social Media Savviness'.

By not including variables on knowledge-rich digital user activity, the indices forego the opportunity to make use of a wealth of digital data, which traces user and activity patterns directly though digital platforms. As such, we maintain that scholars and policy makers interested in knowledge economies would benefit from an index that encourages deeper reflection into what the ranking or score measures and how including variables related to direct user activity changes its interpretation.

Out of the existing indices, the KEI seems to be the index with the highest visibility (Asongu, 2012; Driouchi, Azelmad, & Anders, n.d.; Nguyen & Pham, 2011; Saltelli, 2006; Wielicki & Arendt, 2010). Considering that it was constructed by the World Bank, which itself has a wide operational presence in Sub-Saharan Africa, we assume that the KEI has the strongest influence in how the prospects and challenges related to knowledge economy-related transformations are perceived in Sub-Saharan Africa. As we are interested in integrating measures of digital participation in the evaluation of such prospects, the paper calibrates the KEI with data on digital participation. The remainder of the paper thus proposes the construction of a digital knowledge economy index, which builds on the KEI developed by the World Bank and complements its traditional data sources with bespoke digital data on capacities and skills (measured via content-creation and participation on digital platforms). We first discuss the data in Section 3, proceded by presentation of the index methodology in Section 4, and conclude with a discussion of the ranking and scores of Sub-Saharan Africa on this index in section 5 and the broader implications of these measurement choices in Section 6.

3. Data

The data used in constructing our Digital Knowledge Economy Index (DKEI) are collected on a range of variables, which are aggregated on country level and have yearly observations. The index uses the base of the World Bank Knowledge Economy Index, but adds a fifth subindex that measures digital participation and content creation. The data for the digital participation and content creation subindex is obtained through bespoke methods that allow scraping information directly from websites. The digital participation subindex includes variables measuring collaborative coding activity, local edits to Wikipedia articles, and domain registrations.

Collaborative coding

We maintain that an index measuring the knowledge economy should feature an estimation of programming skills, and data retrieved from GitHub offers an appropriate way of measuring these skills around the world. GitHub is a web-based repository hosting service, which allows users to share and collaborate on software development. Having over 15 million users (GitHub, 2016) in comparison to competitor SourceForge's 3.7 million (SourceForge, 2016) and Launchpad's 3.1 million (Launchpad, 2016), makes GitHub by far the most popular code-hosting service for software development. Measuring collaborative coding provides a way to account for programming skills, which are fundamental for many knowledge-rich activities and industries such as software development.

Coding activity is operationalised by measuring the volume of GitHub 'commits'. Commits are instances of content contribution to GitHub, such as revisions to the code of a software project. The raw commit data that we retrieve through the site's application programming interface (API) is not automatically associated with a country. In order to associate the commit volume with a country, we geocode commits submitted by users that have a parse-able location attribute. We then use the Edina Unlock Places geocoding API to look up country names.

Although we consider the GitHub data to be a suitable proxy for programming skills worldwide, the dataset does have some limitations. Only a quarter of the users indicate their location, and these users account for approximately 45 percent of total commits. However, there is no reason to suspect that a user's choice to indicate their location is subject to significant geographic biases. The process of geocoding user locations is also subject to a margin of error, though recent studies suggest this margin is likely to be narrow (Lima, Rossi, & Musolesi, 2014). We manually reviewed every toponym attached to over 10,000 commits in the dataset, and most popular toponyms per country, which we think renders our geocoding process a good estimate of the user locations. Despite these limitations, we consider GitHub commits to be an appropriate, if imperfect proxy for otherwise hard to measure programming skills.

Wikipedia edits

Wikipedia is the world's leading encyclopedia project, which is written collaboratively by volunteers. Anyone with an Internet access can create or edit Wikipedia articles. Wikipedia is widely considered as one of the largest reference websites and every day tens of thousands of edits and thousands of articles are created to augment the knowledge contained by the encyclopedia (Wikimedia, 2016). We consider the volume of Wikipedia edits to approximate a country's capacity to expand and improve the quality of the knowledge contained in this open resource. The number of Wikipedia edits published from within a country indicates the volume of contributions to the world's largest encyclopaedia. We obtained the data on editing traffic per country as outlined in Graham et al (2015). The dataset is provided by Wikimedia and samples

editing activity across all Wikipedia language versions. The geographic origin of every 1,000th edit is analysed and then added to country-level counts.

While the majority of the contributors to Wikipedia are amateurs writing on non-specialist topics, the fact that content can be edited by anyone tends to result in a consensus of neutral and encyclopedic representation of information and comprehensible, verifiable knowledge. However, some of the editing activity may add false or debatable information or is carried out with the intention of vandalism. Our dataset does not contain metrics on the quality or appropriateness of the Wikipedia edits, but we do not have any reason to suspect that the propensity for vandalism in the editing activity would have a particular geographic bias. Despite these limitations, we regard Wikipedia edits as a unique proxy for digital knowledge creation.

(Domain registrations)

The volume of top-level domains (TLDs) related to a country illustrates the quantity of online content produced around the world. Domain names indicate the volume of codified information and knowledge that is accessible on the web. Evaluating domain registrations offers a unique perspective into countries' online presences (or absences). To operationalize this variable, we sum the three different types of TLDs, including country-code TLDs (ccTLDs) such as .cm for Cameroon, generic TLDs (gTLDs) such as .com or .net, and internationalized TLDs (IDNccTLDs) that are ccTLDs in non-Latin script or alphabet, such as Arabic, or characters such as Chinese.

The processes of associating domain data to countries of registration required different processes for different TLDs. ccTLDs and IDNccTLDs were assigned to their respective countries (eg. all .cm comains were assigned to Cameroon), since previous research characterizes these types of TLDs as emblematic of local content production (Zook, 2001). In order to locate gTLDs, their WHOIS record (record containing the address details of the person or organization registering a domain name) has been geocoded and each domain has been assigned to the respective country.

Yet, the ccTLDs related to certain countries have become extensively popular for being meaningful abbreviations (such as Tuvalu's .tv domain for the entertainment industry), form a word with the domain name (such as bur.st), form a word in a certain language (such as .me), or because they evoke connotations (such as .io being used for start-ups). In an effort to identify upwards-biased ccTLD values we compared the number of ccTLDs with the number of individuals with Internet access for each country. If the number was notably high in countries with low internet penetration, we sought further information regarding the use of the ccTLD and discarded any countries whose ccTLD seemed to be inflated by registrations from outside the associated country (see Table 2 for omitted countries).

Table 2. Countries Omitted Due to Inflated Domain Registrations.

Country	ccTLD	Reason for Omission
Tuvalu	.tv	Used by the media industry
Federated States of Micronesia	.fm	Used by the media industry
Armenia	.am	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 "web site"
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
Tokelau	.tk	The .tk domain can be registered free of charge.
Mali	.ml	The .ml domain can be registered free of charge.
Gabon	.ga	The .ga domain can be registered free of charge.
Central African Republic	.cf	The .cf domain can be registered free of charge.

Since the retained TLDs generally have a clear connection to the address of the person or organization who registered them, we regard the TLD dataset to offer a unique proxy of online content creation.

4. Methodology

The framework and methodology of the World Bank Knowledge Economy Index are used as a basis for our Digital Knowledge Economy Index. We add a fifth subindex measuring digital participation, which carries a weight equal to the other subindices of the KEI. The innovation subindex and the digital participation subindex are included in their weighted forms, as standardizing the score with respect to population size offers a clear estimate of the countries attainment of the variables that are included in these subindices.[iii]

To make the fifth subindex comparable to the existing four subindices, we follow the same normalization procedure as the KEI. Since the digital participation variables are measured in different units and on different scales, we bring the variables to a common standard of measurement through the following procedure:

1. We record the raw data (u) for the digital participation variables.
2. We rank the countries based on their absolute values (rank u). Country with the highest value ranks 1, the second best 2, and so on. Countries with the same value are allocated the same rank.
3. The number of countries ranking higher than a particular country (Nh) is calculated for each country in the sample.
4. The following formula is used to normalize the scores for each country for each variable according to their ranking within the sample and in relation to the total number of countries in the sample (Nc):

   Normalized (u) = 10*(1-Nh/Nc)

The country scores on the fifth subindex are calculated by taking a simple average of the normalized country scores for each of the included variables. The digital knowledge economy score for each country is then calculated as the simple average of the five subindex scores. We gather the scores for each country into an index, which we call the Digital Knowledge Economy Index (DKEI). Based on the DKEI, we also calculate a rank of the included countries, where the country which performs best is ranked one, the second best is ranked two and so forth.

Where a data point is missing for more than one variable within the fifth subindex, we remove the observation in order to maintain data integrity. This is in line with the KEI methodology, which allows data to be missing for at most one variable per subindex.[iv]

The digital participation variables are measured for year 2013, while the variables included in the KEI are measured for year 2012. While we do not suspect that the one year gap between the measures is large enough to warrant further corrective measures, we monitor the relationship of the fifth subindex to the other subindices closely throughout the analysis.

5. Results

To facilitate discussion of the results, we calculated differences for the change in the ranks between the DKEI and KEI as well as for the change between the DKEI and KEI scores (see Table 3).

Table 3. Digital Knowledge Economy Index (DKEI) 2012 Rankings.

Country	DKEI 2012 Rank	KEI 2012 Rank	Rank Change	DKEI 2012 Score	KEI 2012 Score	Score Change
Sweden	1	1	0	9.44	9.43	0.01
Finland	2	2	0	9.27	9.33	-0.06
Denmark	3	3	0	9.21	9.16	0.05
Norway	4	4	0	9.19	9.11	0.08
Netherlands	5	5	0	9.15	9.11	0.04
New Zealand	6	6	0	9.09	8.97	0.12
Canada	7	7	0	9.02	8.92	0.10
Australia	8	9	1	9.00	8.88	0.12
Switzerland	9	10	1	8.96	8.87	0.09
Ireland	10	11	1	8.96	8.86	0.10
United Kingdom	11	14	3	8.94	8.76	0.18
United States	12	12	0	8.92	8.77	0.15
Germany	13	9	-4	8.84	8.9	-0.06
Iceland	14	16	2	8.82	8.62	0.20
Belgium	15	15	0	8.69	8.71	-0.02
Austria	16	17	1	8.56	8.61	-0.05
Estonia	17	19	2	8.50	8.4	0.10
Luxembourg	18	20	2	8.42	8.37	0.05
Taiwan	19	13	-6	8.33	8.77	-0.44
Spain	20	21	1	8.23	8.35	-0.12
Singapore	21	23	2	8.22	8.26	-0.04

Czech Republic	22	26	4	8.20	8.14	0.06
Slovenia	23	28	5	8.17	8.01	0.16
Israel	24	25	1	8.15	8.14	0.01
France	25	24	-1	8.14	8.21	-0.07
Hungary	26	27	1	8.03	8.02	0.01
Hong Kong	27	18	-9	8.02	8.52	-0.50
Japan	28	22	-6	7.97	8.28	-0.31
Lithuania	29	32	3	7.93	7.8	0.13
Malta	30	32	1	7.87	7.88	-0.01
Italy	31	30	-1	7.78	7.89	-0.11
Portugal	32	34	2	7.75	7.61	0.14
Cyprus	33	35	2	7.65	7.56	0.09
Slovak Republic	34	33	-1	7.61	7.64	-0.03
Greece	35	36	1	7.59	7.51	0.08
South Korea	36	29	7	7.59	7.97	-0.38
Latvia	37	37	0	7.54	7.41	0.13
Croatia	38	39	1	7.47	7.29	0.18
Poland	39	38	-1	7.44	7.41	0.03
Barbados	40	41	1	7.04	7.18	-0.14
Bulgaria	41	45	4	7.04	6.8	0.24
Chile	42	40	-2	6.99	7.21	-0.22
Romania	43	44	1	6.97	6.82	0.15
United Arab Emirates	44	42	-2	6.74	6.94	-0.20
Bahrain	45	43	-2	6.63	6.9	-0.27
Uruguay	46	46	0	6.45	6.39	0.06
Serbia	47	49	2	6.26	6.02	0.24
Malaysia	48	48	0	6.16	6.1	0.06
Costa Rica	49	51	2	5.98	5.93	0.05
Trinidad and Tobago	50	52	2	5.97	5.91	0.06
Qatar	51	53	2	5.92	5.84	0.08
Russian Federation	52	54	2	5.91	5.78	0.13
Macedonia	53	56	3	5.82	5.65	0.17
Ukraine	54	55	1	5.80	5.73	0.07
Belarus	55	58	3	5.68	5.59	0.09
Argentina	56	61	5	5.66	5.43	0.23
Brazil	57	59	2	5.62	5.58	0.04
Mauritius	58	60	2	5.61	5.52	0.09
Saudi Arabia	59	50	-9	5.51	5.96	-0.45
Oman	60	47	-13	5.49	6.14	-0.65
Jamaica	61	57	-4	5.49	5.65	-0.16
Panama	62	63	1	5.45	5.3	0.15
Armenia	63	69	6	5.44	5.08	0.36
South Africa	64	64	0	5.41	5.21	0.20
Turkey	65	67	2	5.35	5.16	0.19
Georgia	66	66	0	5.32	5.19	0.13
Bosnia and Herzegovina	67	68	1	5.32	5.12	0.20
Kuwait	68	62	-6	5.19	5.33	-0.14

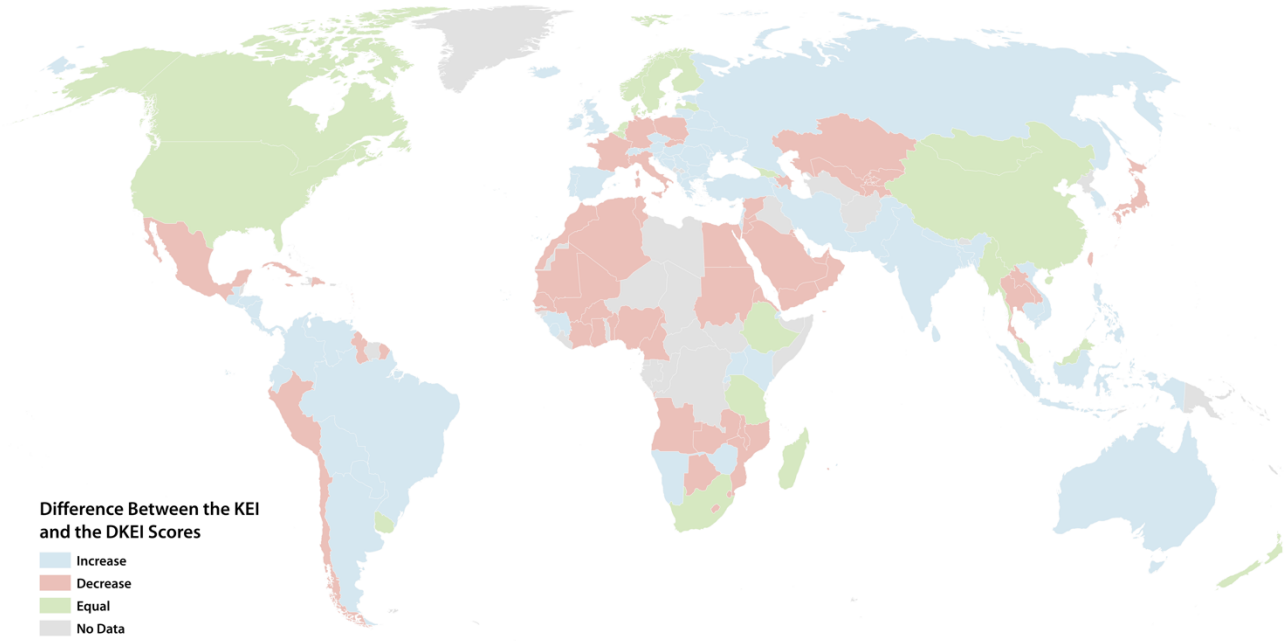
Moldova	69	75	6	5.17	4.92	0.25
Thailand	70	65	-5	5.13	5.21	-0.08
Mexico	71	70	-1	5.10	5.07	0.03
Colombia	72	74	2	5.08	4.94	0.14
Peru	73	72	-1	4.96	5.01	-0.05
Jordan	74	73	-1	4.90	4.95	-0.05
Lebanon	75	79	4	4.87	4.56	0.31
Kazakhstan	76	71	-5	4.77	5.04	-0.27
Albania	77	80	3	4.70	4.53	0.17
Azerbaijan	78	77	-1	4.50	4.56	-0.06
Guyana	79	76	-3	4.41	4.67	-0.26
Venezuela	80	84	4	4.36	4.2	0.16
Mongolia	81	81	0	4.34	4.42	-0.08
China	82	82	0	4.33	4.37	-0.04
Tunisia	83	78	-5	4.28	4.56	-0.28
El Salvador	84	86	2	4.25	4.17	0.08
Philippines	85	91	6	4.16	3.94	0.22
Namibia	86	87	1	4.15	4.1	0.05
Paraguay	87	89	2	4.02	3.95	0.07
Iran	88	92	4	4.02	3.91	0.11
Botswana	89	83	-6	4.02	4.31	-0.29
Fiji	90	90	0	4.00	3.94	0.06
Dominican Republic	91	88	-3	3.94	4.05	-0.11
Sri Lanka	92	99	7	3.91	3.63	0.28
Ecuador	93	96	3	3.79	3.72	0.07
Cuba	94	85	-9	3.78	4.19	-0.41
Bolivia	95	98	3	3.68	3.68	0.00
Guatemala	96	97	1	3.67	3.7	-0.03
Kyrgyz Republic	97	93	-4	3.63	3.82	-0.19
Egypt	98	95	-3	3.62	3.78	-0.16
Cabo Verde	99	101	2	3.60	3.59	0.01
Vietnam	100	102	2	3.56	3.4	0.16
Morocco	101	100	-1	3.53	3.61	-0.08
Algeria	102	94	-8	3.46	3.79	-0.33
India	103	108	5	3.26	3.06	0.20
Honduras	104	107	3	3.15	3.08	0.07
Indonesia	105	106	1	3.12	3.11	0.01
Kenya	106	109	3	2.91	2.88	0.03
Nicaragua	107	113	6	2.86	2.61	0.25
Swaziland	108	104	-4	2.85	3.13	-0.28
Uzbekistan	109	103	-6	2.82	3.14	-0.32
Tajikistan	110	105	-5	2.79	3.13	-0.34
Syrian Arab Republic	111	110	-1	2.73	2.77	-0.04
Ghana	112	111	-1	2.71	2.72	-0.01
Pakistan	113	115	2	2.60	2.45	0.15
Senegal	114	112	-2	2.53	2.7	-0.17
Uganda	115	116	1	2.41	2.37	0.04

Zambia	116	114	-2	2.22	2.56	-0.34
Zimbabwe	117	118	1	2.13	2.17	-0.04
Nigeria	118	117	-1	2.09	2.2	-0.11
Rwanda	119	125	6	2.00	1.83	0.17
Nepal	120	133	13	1.95	1.58	0.37
Cambodia	121	130	9	1.86	1.71	0.15
Malawi	122	120	-2	1.83	1.92	-0.09
Yemen	123	121	-2	1.78	1.92	-0.14
Lesotho	124	119	-5	1.78	1.95	-0.17
Mali	125	124	-1	1.77	1.86	-0.09
Tanzania	126	126	0	1.75	1.79	-0.04
Madagascar	127	127	0	1.74	1.77	-0.03
Benin	128	123	-5	1.71	1.88	-0.17
Burkina Faso	129	122	-7	1.70	1.91	-0.21
Laos	130	129	-1	1.70	1.75	-0.05
Bangladesh	131	135	4	1.68	1.49	0.19
Mozambique	132	128	-4	1.58	1.76	-0.18
Cameroon	133	131	-2	1.56	1.69	-0.13
Mauritania	134	132	-2	1.42	1.65	-0.23
Cote d'Ivoire	135	134	-1	1.39	1.54	-0.15
Djibouti	136	137	1	1.37	1.34	0.03
Sudan	137	136	-1	1.28	1.48	-0.20
Ethiopia	138	138	0	1.15	1.27	-0.12
Eritrea	139	140	-1	1.10	1.14	-0.04
Guinea	140	139	1	1.04	1.22	-0.18
Sierra Leone	141	142	1	1.02	0.97	0.05
Angola	142	141	-1	1.02	1.08	-0.06
Myanmar	143	143	0	1.02	0.96	0.06

Note: The cell indicating rank change is highlighted in blue for countries that rank higher in the DKEI than in the KEI, red for countries that rank lower in the DKEI than in the KEI, and green for countries that rank the same. The entire row is highlighted for Sub-Saharan African countries.

Overall, the ranking of the DKEI resembles the KEI, with seven of the highest ranking countries on KEI maintaining their ranks on DKEI. However, the countries that rank in the upper and middle ranges of the DKEI have to a large extent improved from their KEI ranking. On the contrary, countries that rank in the lowest ranges of the DKEI, tend to rank worse than they do in the KEI. The majority of Sub-Saharan African countries belong to this final group (see figure 2). While Mauritius and South Africa continue to rank higher than the rest of the group in the DKEI, the ranking of most of the Sub-Saharan African countries declines when the digital participation subindex is included in the ranking.

Figure 2. The Geography of the Differences Between the KEI and the DKEI Scores.



For the Sub-Saharan African countries whose ranking is lower in the DKEI than in the KEI, the average change by which the ranking has declined is 2.5 ranks. While the change in rank of most of the countries in this group is around the average drop, the ranking of a few countries dipped substantially more, including Burkina Faso (-7), Botswana (-6), Benin (-5), Lesotho (-5), Swaziland (-4), and Mozambique (-4). Some of the continent's countries rank higher in the DKEI than in the KEI, and the average change by which the ranking has increased within this group is 2 ranks. Notable deviations from these averages include Rwanda and Kenya, which improved by 6 and 3 ranks, respectively.[v]

As the final column of the table indicates, the scores of the DKEI and KEI show only limited variation between them. Observing the change in scores for Nepal, which moved thirteen ranks between the two indices, the difference between the index scores is 23 percent of the KEI value. The average increase in the score within the Sub-Saharan countries that experienced a jump up is 0.07, whereas the average drop between the countries that scored worse on DKEI than KEI is 0.16.

Given that the digital participation subindex carries the same weight as the four KEI subindices, it makes up 20 percent of the DKEI score, whereas the respective impact of the KEI is 80 percent. Thus limited changes in the actual index scores are not surprising. However, recalibrating the KEI with the digital participation subindex urges closer attention to the appropriateness of the KEI normalization procedure.

Although normalization is helpful in order to make the scales of the variables comparable, the degree of such standardizing measure matters as well. For instance, while it is important that variables with positive (where higher values mean better performance) and negative (where lower values mean better performance) directions are anchored onto one common direction, the extent to which standardization alters the relation between the values of those variables will intimately affect how the index compares the performance of countries. The KEI normalizes the actual variable scores into a score that takes a value between 0 and 10, which unifies their direction and sets a common upper and lower bound. However, the procedure allocates these normalized scores not based on the variable scores, but based on the ranking of the variable

scores. As a result, the countries' DKEI and KEI scores are distributed evenly along the upper and lower bound of the index. The top performing ten percent have a normalized index score between 10 and 9, whereas the lowest performing countries have a score between 0 and 1. While this normalization procedure helps to rank the countries in terms of their performance, it smooths over the differences between their performance. Given that digital participation as well as in various other variables included in the KEI are characterized by skewed distributions between countries, even where they have been standardized by population, the KEI standardization method might not adequately capture their variation and thus render the KEI and DKEI scores biased with respect to this distribution. An alternative standardization procedure such as the min-max method could ameliorate this issue.[vi] In the min-max method, each variable is projected onto a scale between 0 and 1, where the minimum value in the variable's scale is anchored to value 0 on the normalized scale, and the maximum value in the variable's scale is anchored to value 1 on the normalized scale. This normalization method would achieve a unified minima and maxima for the index scores, but would allow their distribution to retain the distribution of the actual variable scores.

6. Discussion and Conclusion

Re-estimating the KEI with the inclusion of a digital participation subindex tells a story about Sub-Saharan Africa, where expectations about digital development and the information age are not met with equal enthusiasm in terms of participation through digital platforms and avenues. The DKEI rankings saw two thirds of countries in Sub-Saharan Africa fall in comparison to their ranking in the KEI, and the rate of this drop was higher on average than the the increase in the ranking of those few countries that improved their KEI rank in the DKEI. Although a few countries in the region (such as Rwanda and Kenya) have paid considerable attention to digitalization, their improved DKEI scores are an outlier within the larger group. For the majority of Sub-Saharan countries, including a measure of digital participation in the estimation of their attainment or preparedness for knowledge economy transformation seems to indicate challenges rather than prospects. This is a sobering reminder for policy and business circles, where knowledge economy visions are fuelled with hope and hype about the leapfrogging prospects of digitalization.

On a more practical level, the re-calibration of the KEI with a digital participation subindex raises important questions about the appropriateness of the KEI normalization procedure for the purpose that the index is meant to serve. This is especially important given Given its influence in shaping how the prospects and challenges related to knowledge economy transformations are perceived (and, as a result, valued and acted upon) by actors working in Sub-Saharan Africa. By smoothing skewed distributions and standardizing variance, the KEI methodology fails to reveal where the underlying data would show both wide and narrow gaps between ranks rather than nearly uniform differences. This is critical information to those in charge of decisions about funding and implementation and could help them to design feasible policies that have a better chance at success. Indices such as the DKEI offer a useful tool for policymakers and other actors in the region. But it could be even more advantageous by allowing index scores to reflect the distances between the values of the underpinning data. Further, in order to allow more granularity, index scores could be re-estimated exclusively for Sub-Saharan countries or within a certain income group to allow for a comparison that might offer more relvant insight than an index with a global coverage.

While including the data on digital participation offers novel insights into Sub-Saharan Africa's efforts to build knowledge economies, these data are not without their limitations. GitHub and Wikipedia do not enforce their users to indicate a country location, which renders some user activity challenging to locate. Domain registrations are subject to a variety of registration processes across the world, which makes it diffuclt to estimate the proportion of websites created from within countries whose domain names are popular with foreign users. Although these limits require care in the analysis and interpretation of digital participation, the variables used remain important proxies and likely the best available measures for otherwise difficult to measure characteristics of the knowledge economy. Despite their limitations, these variables offer a unique perspective into digital participation, which we view as critical in exposing patterns of contemporary skills and knowledge that should be included in estimation of knowledge economy attainment.

This exercise has demonstrated that indices evaluating knowledge economy are far from being neutral or value free technical tools, and that non-neutral choices are made on every step of creating such an index. While the limits and decisions pertaining to some of these choices are discussed rather transparently in the material accompanying the index, various other choices such as rationale of including certain variables, construction of the methodology and appropriateness of normalization procedure are rarely discussed in great detail, and nowehere are accompanied by reflection of the structures of the underlying data. Overly detailed technical discussions or data analysis may not feasible in the fast-paced policy, business and donor circles, but we hope that easy-to-use tools such as indices would be accompanied with more grounded information about measurement choices. Complex issues and phenomena such as 'knowledge economy' are challenging to measure in a single number, and none of the indices discussed in this study aim to offer an all-encompassing analysis. However, given their userfriendliness, indices are quick in influencing perceptions especially in an environment where decisions are made quickly or with limited resources for deliberation. Given their tendency to offer a handy heuristic device, the indices illustrating complex phenomena should not be published without sufficient information about the underlying measurement choices. Further, we hope that academics working in the field of international development will be active in publishing studies of indices measuring critical aspects of development, and provide the reflexive tone where that is missing.

Funding

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

Notes

[i] When discussed in a context of national development, the concept of leapfrogging tends to be associated with forgoing investment in agricultural-intensive economies and labor-intensive economies, and even service-based economies and focusing more directly on knowledge-based industries, where knowledge resources such as trade secrets, brands and expertise are as critical as other economic resources.

[ii] We thought of comparing the Sub-Saharan African regional average to the European regional average, but as the inclusion of countries from both regions varies between the indices, we decided to compare the Sub-Saharan regional average to the United States, as it is a well-established knowledge economy with a large population and thus makes sense to consider as a single country.

[iii] We think that in the contect of measuring and comparing countries' overall knowledge economy attainment, it is helpful to standardize the variables by the population. This weighted form of the KEI is also the default index configuration offered by the World Bank. However, we recognize that the unweighted variables measuring total numbers of digital participation and innovation tell an important story as well, since absolute size of resources matters where a critical mass of creativity and innovation is needed in order to facilitate exchange of ideas in a certain location. Further, populous economies such as China and India have a critical mass of innovative capacity, which is reflected less prominently when variables are scaled by population. However, as the Sub-Saharan African countries are the main focus of this study, we chose to use the weighted versions of both KEI and DKEI.

[iv] Because of this requirement, we removed Dominica and Aruba from the dataset and recalculated the rankings for the 143 variables included in the KEI.

[v] Given that the rate of digital participation across the three digital variables shows a growing time trend, and that the DKEI is estimated with data from 2013, these estimates are likely slightly upwards biased. If the data from 2012 were available for the digital subindex, the countries that experienced lower DKEI scores than KEI scores, would likely have even lower DKEI scores, and the countries that had higher DKEI scores than KEI scores might have had lower increase in their scores, or perhaps no increases at all.

[vi] While it would have been interesting to re-estimate the KEI and DKEI using the min-max method, unfortunately the World Bank does not release the non-normalized dataset on which the KEI was estimated.

Bibliography

Asongu, S. A. (2012). Financial Sector Competition and Knowledge Economy: Evidence from SSA and MENA Countries. *Journal of the Knowledge Economy*, *6*(4), 717–748. http://doi.org/10.1007/s13132-012-0141-4

Baltowski, M., & Pastuszak, Z. (2008). Sources of the success of Scandinavian knowledge economies. *International Journal of Innovation and Learning*, *5*(2), 109–118. http://doi.org/10.1504/IJIL.2008.016759

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. http://doi.org/10.1080/02681102.2012.719859

Chen, D. H. C., & Dahlman, C. J. (2006). The Knowledge Economy, the KAM Methodology and World Bank Operations. The International Bank for Reconstruction and Development/The World Bank. Retrieved from http://documents.worldbank.org/curated/en/695211468153873436/pdf/358670WBI0The11dge1Economy01PUBLIC1.pdf

Devarajan, S. (2013). Africa's Statistical Tragedy. *Review of Income and Wealth*, *59*, S9–S15. http://doi.org/10.1111/roiw.12013

Driouchi, A., Azelmad, E. M., & Anders, G. C. (n.d.). An Econometric Analysis of the Role of Knowledge in Economic Performance. *The Journal of Technology Transfer*, *31*(2), 241–255. http://doi.org/10.1007/s10961-005-6109-9

Friederici, N., Ojanperä, S., & Graham, M. (2016). *The Impact of Connectivity in Africa: Grand Visions and The Mirage of Inclusive Digital Development*. Manuscript submitted for publication.

GitHub. (2016). Build software better, together. Retrieved August 27, 2016, from https://github.com

Graham, M., Andersen, C., and Mann, L. 2015 Geographical Imagination and Technological Connectivity in East Africa. *Transactions of the Institute of British Geographers* 40(3) 334-349.

Graham, M., Straumann, R. K., & Hogan, B. (2015). Digital Divisions of Labor and Informational Magnetism: Mapping Participation in Wikipedia. *Annals of the Association of American Geographers*, *105*(6), 1158–1178. http://doi.org/10.1080/00045608.2015.1072791

Huggins, R. (2008). The evolution of knowledge clusters: Progress and policy. *Economic Development Quarterly*, *22*(4), 277–289. http://doi.org/10.1177/0891242408323196

Jerven, M. (2013a). Comparability of GDP estimates in Sub-Saharan Africa: The effect of Revisions in Sources and Methods Since Structural Adjustment. *Review of Income and Wealth*, *59*, S16–S36. http://doi.org/10.1111/roiw.12006

Jerven, M. (2013b). *Poor numbers: how we are misled by African development statistics and what to do about it*. Ithaca: Cornell University Press.

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.

Launchpad. (2016). People and teams in Launchpad. Retrieved August 27, 2016, from https://launchpad.net/people

Lehohla, P. (2008). Statistical development in Africa in the context of the global statistical system. *Statistical Journal of the IAOS*, *25*(1-2), 27–45.

Lima, A., Rossi, L., & Musolesi, M. (2014). Coding Together at Scale: GitHub as a Collaborative Social Network. *arXiv Preprint arXiv:1407.2535*.

Nguyen, T. V., & Pham, L. T. (2011). Scientific output and its relationship to knowledge economy: An analysis of ASEAN countries. *Scientometrics*, *89*(1), 107–117. http://doi.org/10.1007/s11192-011-0446-2

Piaggesi, D., & Chea, M. J. (2011). The knowledge economy: A new development paradigm for Latin America and the Caribbean (LAC). In *Global Strategy and Practice of E-Governance: Examples from Around the World* (pp. 464–477).

Saltelli, A. (2006). Composite Indicators between Analysis and Advocacy. *Social Indicators Research*, *81*(1), 65–77. http://doi.org/10.1007/s11205-006-0024-9

SourceForge. (2016). About. Retrieved August 27, 2016, from https://sourceforge.net/about

The World Bank. (2013). Knowledge Assesment Methodology 2012. Retrieved from www.worldbank.org/kam

Wielicki, T., & Arendt, L. (2010). A knowledge-driven shift in perception of ICT implementation barriers: Comparative study of US and European SMEs. *Journal of Information Science*, *36*(2), 162–174. http://doi.org/10.1177/0165551509354417

Wikimedia. (2016). Wikipedia:About. In *Wikipedia, the free encyclopedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Wikipedia:About&oldid=735677110

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. http://doi.org/10.1177/00027640121958113