



Partnership for FINANCE
in a DIGITAL AFRICA

Focus note

Can Big Data Shape Financial Services in East Africa?

ACKNOWLEDGEMENTS

The author of this Focus Note is Marissa Dean, based on research conducted in Kenya and Tanzania by Marissa and Caribou Digital colleague Annabel Schiff.

This research was supported by the Mastercard Foundation, and we are grateful to Youssouf Sy and Mark Wensley for their support and comprehensive input. We thank our colleagues from Caribou Digital David Edelstein, Jonathan Donner, Maha Khan, and Niamh Barry for comments and inputs which significantly improved this document.

We also thank all of the organizations who participated in the research for their valuable insights, as well as Accion Venture Labs, Bankable Frontiers Associates, CGAP, GSM, Mercy Corps, and UNCDF colleagues who provided helpful introductions and feedback.

NOTES

The views presented in this paper are those of the author and the Partnership, and do not necessarily represent the views of the Mastercard Foundation or Caribou Digital.

For questions or comments please contact us at ideas@financedigitalafrica.org.

RECOMMENDED CITATION

Partnership for Finance in a Digital Africa, "Focus Note: Can Big Data Shape Financial Services in East Africa?" Farnham, Surrey, United Kingdom: Caribou Digital Publishing, 2018.
<http://www.financedigitalafrica.org/research/2018/03/can-big-data-shape-financial-services-in-east-africa/>.

ABOUT THE PARTNERSHIP

The Mastercard Foundation Partnership for Finance in a Digital Africa (the "Partnership"), an initiative of the Foundation's Financial Inclusion Program, catalyzes knowledge and insights to promote meaningful financial inclusion in an increasingly digital world. Led and hosted by Caribou Digital, the Partnership works closely with leading organizations and companies across the digital finance space. By aggregating and synthesizing knowledge, conducting research to address key gaps, and identifying implications for the diverse actors working in the space, the Partnership strives to inform decisions with facts, and to accelerate meaningful financial inclusion for people across sub-Saharan Africa.

www.financedigitalafrica.org

www.mastercardfdn.org/financial-inclusion-program/



This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>.

Readers are encouraged to reproduce material from the Partnership for Finance in a Digital Africa for their own publications, as long as they are not being sold commercially. We request due acknowledgment, and, if possible, a copy of the publication. For online use, we ask readers to link to the original resource on the www.financedigitalafrica.org website.

Contents

1	Executive summary	1
2	Insights	4
	Limited use case for third-party data	5
	Organizations are pursuing partnerships in lieu of transactional relationships	5
	Most of the organizations interviewed are still testing, refining, and experimenting with data analytics	10
	Most banks and FinTechs believe a business case that is strategy- and leadership-led is essential to the uptake of big data and analytics	12
3	Conclusions	14
4	Appendix	16
	Bibliography	20

1 Executive summary

The volume of digital data created by users on the African continent is growing. With this growth come new opportunities for alternative “big data”¹ to catalyze an expansion of financial services to low-income and hard to reach populations. M-Shwari, the high profile micro-savings and loan product developed by Safaricom and Commercial Bank Africa (CBA), has already achieved impressive volumes by capitalizing on this trend.² Moreover, M-Shwari’s success demonstrates a business case for banks to leverage big data—in the form of mobile phone records and mobile money behavior—to expand low denomination lending.³

Indeed, mobile credit services are leveraging usage and behavior data to determine which customers can be granted small amounts of credit (cash or airtime) over their mobile phone. However, financial services organizations in East Africa are still exploring whether there is value in mobile data and other big datasets, such as business operations data, social media discussions, and e-commerce digital trails. More data can mean more problems: there is complexity in integrating unstructured, non-traditional data with existing practices on both the technical level and the strategic level. In addition to demonstrating that lending to risky, low-income segments generates sufficient returns, organizations that want to use big data to reach new customers must overcome substantial transformation hurdles—including IT costs, business culture, and process changes.

FinTech startups operating as alternative lenders⁴ have fewer legacy technology constraints, but they face challenges in acquiring low-cost wholesale capital, especially for FinTech startups lending to sectors which are perceived by banks as high risk.⁵ These FinTechs must also deal with the issue of whether big data can serve as an adequate predictor of risk and whether the business itself can generate sufficient profits by servicing low-income or difficult to reach end-users with digital financial services.

This Focus Note summarizes how 30 leading organizations⁶ across Kenya and Tanzania are thinking about or actively using big data and analytics as they explore this space in digital finance. The findings are based on interviews with these organizations—which include banks, microfinance institutions (MFIs), mobile network operators (MNOs), and FinTech organizations (listed in [Appendix: Research Approach](#))—conducted between August and September 2017.

The interviews reveal that while players in East Africa do not exchange and profit from big data directly, there is a growing demand for traditional and big datasets. This has led to a slow and steady uptake of big data and analytics. The findings further indicate there are four key factors driving the use of big data analytics, discussed in detail in the next section. The findings thus specifically focus on the drivers affecting the use of big data and analytics as well as the potential for big data and analytics to enable organizations to better reach low-income and difficult to reach customers.

1 “Alternative data” is information gathered from sources not traditionally used for underwriting purposes. Increases in digital data trails from smartphones, computers, sensors, and other devices has led to large amounts of structured, semistructured, and unstructured “big data” that has the potential to be used in underwriting processes. These two concepts overlap but are not fully interchangeable. Most of the data discussed in this Focus Note can be considered both “big data” and “alternative data.” Some data, however, is either alternative or traditional but not “big data.” For simplicity, this Focus Note will use “big data” to denote “big, alternative datasets” and “traditional” for traditional underwriting data. For a more detailed description and characteristics of “big data,” see FiDA’s Snapshot 9, “[Best practices in big data analytics](#).”

2 As discussed in CGAP’s [How M-Shwari Works: The Story So Far](#), by the end of 2014, M-Shwari had disbursed KES 20.6 million in loans to 2.8 million borrowers, with a non-performing loan rate of 2.2% (after 90 days).

3 See FiDA’s Snapshot 8: “[What is the commercial landscape of digital finance?](#)” for more details on M-Shwari’s ROI and other metrics.

4 Alternative lenders are non-banking institutions using digitally based platforms to provide a broad range of loan products to consumers, SMEs, etc.

5 See FIBR’s briefing note [Alternative lending: landscaping the funding models for lending fintech companies](#) for more background on constraints alternative lenders face lending on balance sheet.

6 For a list of organizations interviewed, see [Appendix: Research Approach](#).

This Focus Note is primarily for organizations currently exploring or investing in big data and analytics technology in East Africa in order to implement new projects as a key component of strategies to reach new customers. It is also useful for research and policy-making organizations striving to advance financial inclusion and looking to understand the current state of big data and analytics usage in East Africa. We trust that the reflections and recommendations presented in this Focus Note will help organizations take informed action with regard to the use of big data and analytics.

2 Insights

A market allowing players to profit directly by exchanging big data is lacking in East Africa, but there is emerging demand for analytics of traditional or big datasets. Very few organizations pay or get paid directly for any kind of data aside from credit reference bureaus (CRBs—see [Box 1](#) for more detail from interviewees about the role CRBs are playing in Kenya).

One of the biggest misconceptions in this industry is that data today has real world commercial value. Data has internal operational value certainly, but almost no one is purchasing data. There isn't enough data, it isn't standardized, and no one has the money to pay for it.

— Lendable

Selling data directly isn't very valuable. Selling insights from that data has value.

— FarmDrive

As noted above four main factors, discussed in detail below, currently drive the use of big data and analytics:

- 1 Lenders have a limited use case for third-party data.
 - 2 Organizations are pursuing partnership models in lieu of transactional relationships.
 - 3 Most organizations are still testing, refining, and experimenting with data analytics.
 - 4 It is critical for organizations to develop a strategy and gain leadership support to employ big data and analytics.
-
-

Limited use case for third-party data

When generating an overall risk rating for a customer, financial services organizations utilize defined risk rating methodologies, such as a credit-scoring model, that ultimately calculates a summary score based on a select list of variables. The two factors that contribute most to credit scoring are *ability to pay* and *willingness to pay*.⁷

- An *ability to pay* score rates the customer's capacity to take on additional debt. In traditional underwriting, this reflects an analysis of account balances, sources of income, and current amounts owed, among other factors.
- A *willingness to pay* score rates the customer's willingness to repay the additional debt. In traditional underwriting this is based on payments history (e.g., how often the customer pays her accounts on time), but, effectively, this is a score of the applicant's character.⁸

In order to determine *ability to pay*, interviews with lenders indicated that the information captured directly by organizations (internal data) is most valuable. Traditional underwriting practices still dominate credit decisions and repayment data is seen as the most powerful predictor of risk. Indeed, Equity Bank mentioned that, according to their sensitivity analysis, their internal data score (based on cash-in-cash-out transactions, savings account information, and current account information) is a more powerful predictor of risk than a credit reference bureau score or a psychometric score. Moreover, some of the FinTechs interviewed shared this perspective, suggesting that

organizations only use external big data if internal data on repayments is unavailable.

The value of external data relative to internal data is only high if internal data is pretty weak. If you have weak internal data, it's the result of weak processes and weak compliance.

— First Access

Nothing tells you better about how someone will pay than how they have paid in the past.

— Lendable

When big data is used to determine ability to pay, the predominant use case is to make an initial credit decision for a low denomination, short-term loan. As illustrated in **Box 2**, this scenario is common in cases in which MNO data is leveraged to set an initial credit limit for a new client.

Big data, however, is proving increasingly useful in assessing *willingness to pay* once a prospective client's ability to pay has been established. Banks and MFIs indicated that they have already experimented with or are currently testing models with FinTechs that provide big data analytics that can be used for this purpose.

There is a potential cost savings if data and analytics prove to be both less expensive and as good or better than human touch (e.g., field officers) at assessing risk. Interviews with lenders suggest that lenders have divergent views regarding the extent to which processes can be entirely digital. Equity Bank has been working to become process-light and cash-light as well as to enable customers to

⁷ For example in Juhudi Kilimo's credit scoring model, Ability and Character combined contribute 80% of the total score.

⁸ *Ability to pay* and *willingness to pay* are two categories of input that financial institutions use to evaluate credit risk. As discussed in this Focus Note, combining multiple sources of input data is the most effective way to evaluate credit worthiness, particularly for smallholder farmers who rely on agricultural livelihoods. According to Mercy Corps AgriFin Accelerate's paper, "[Digital financial services for smallholder farmers: what data can financial institutions bank on?](#)", while a farmer's history of repaying loans is strongly related to willingness to pay, it does not offer insight into a farmer's ability to plant, harvest and sell particular crops, or to breed, raise and generate income from animals. Thus, the value of using loan repayment histories in credit scoring models is likely a function of multiple factors, such as the cost of collecting the data, the extent to which the data is available from all applicants, accuracy of the data, and the relevance of the data to a farmer's ability or willingness to repay a loan. Moreover, there are diminishing returns to using more than ten or so factors in credit risk scoring models for smallholders.

engage directly with the bank without having to rely on human interfaces. Currently 93% of loans are facilitated through mobile phones. Once integration with national population registries is complete, the bank anticipates that 95% of the account registration process will be able to be done via mobile phones. FINCA Impact Finance is similarly working toward making their business entirely digital. The organization is currently piloting digital data underwriting in several branches and will be largely moving out of traditional underwriting in 2018. Apollo Agriculture also believes that digital can reduce the cost of lending to farmers—by combining machine learning with satellite data, they have created a process to make credit decisions and communication with farmers through mobile phones without relying on field staff. In contrast, Musoni Kenya is 100% cashless and nearly paperless, but still relies on traditional underwriting practices with a very high-touch, subjective process.

Psychometric big data—including online quizzes to judge character or personality traits and analysis of Facebook “likes”—is garnering increased attention. Suppliers of psychometric data or psychometric tools, such as EFL, believe not only that their data and analytics are predictive but also that they have a key advantage in their applicability to everyone, even clients with limited credit history (“thin-file” clients), as a starting point. When layered with other big and traditional data sources (e.g., social media, mobile phone, bureau data, bank historical data), proponents expect psychometrics to become even more powerful. Indeed, Equity Bank conducted an experiment with EFL’s psychometric scoring model and found it both predictive and useful; they plan to integrate it into applicable models across their regional subsidiaries.⁹ Moreover, Juhudi Kilimo decided to partner with EFL in order to evaluate character as part of their risk assessment. This was previously carried out by loan officers, but they believed the EFL approach would be more objective.

However, lenders that have experimented with psychometrics, as well as other self-reported data, have found some limitations. Validating quality and authenticity can be challenging—some things can be validated by satellite but others simply cannot. Apollo Agriculture mentioned that although they capture self-reported data from farmers, it can be

difficult to get a full profit and loss statement from an individual farmer. This could be because, as noted by one FinTech, farmers modify responses depending on why they think the information is being collected. As a result, providers try not to rely too heavily on self-reported data. Moreover, when data comes from a third party, providers often do not validate the identities of respondents. This creates an opportunity for people to game the system. Indeed, Tala indicated that consumers in Kenya are extremely savvy; because digital lenders tell applicants why they have been declined, customers simply reapply and respond differently.¹⁰ Additionally, Arifu and Juhudi Kilimo both noted that it is unclear whether a client might have passed the phone to a family member to take a test. Additionally, another digital lender relying on mobile operator KYC data encountered an instance of fraud wherein more than 100 loans were taken out by a fraud ring using the same name of a popular fictitious character for all the SIM card registrations.

In addition to individualized datasets, organizations also use marketwide data, such as agricultural data, household income levels, and crop prices. Most marketwide datasets currently available have zero value for determining an individual’s ability to pay and willingness to pay, and limited value for segmentation. However, some datasets can be used for back modelling.¹¹ Government data and other publicly available datasets have huge variations making them difficult to use for any trending because organizations cannot discern which data to trust when the margin of error among sources is so wide.¹² For instance, Off Grid Electric mentioned that publicly-available data sources on electrification rates range from 10–92% in some regions. Further validation or spot checking by a human is necessary, thereby nullifying most of the utility of having a dataset in the first place. The data itself is neither structured nor sufficiently detailed to be segmented in a meaningful way, rendering it practically useless beyond gaining a very general picture. Moreover, the task of integrating anonymized data with individual data is incredibly difficult and not something most organizations are equipped to handle.

One specific type of marketwide data—satellite data—is proving useful for several FinTechs, and appears to be a relatively untapped opportunity. For organizations that can make use of it, there are clear

9 As discussed in subsection 4, the interest rate cap in Kenya makes lending to risky segments challenging because there is no wiggle room to price in risk. Likewise, there is little room to price in alternative credit scoring methods.

10 While some organizations noted that customers may change their loan application details to obtain a favorable loan, there is a strong movement toward strengthening providers’ terms and conditions in order to protect customers. For example, the SMART Campaign, advocates the use of [Client Protection Principles](#) in offering nano-loans through mobile phones. In their paper “[Tiny Loans Big Questions](#),” SMART emphasizes the need for the digital finance community to define and implement mobile credit consumer protection best practices in order to prevent damaging consumer trust, regulatory clampdowns, and over-indebtedness crises.

11 Back modelling or backtesting means testing a predictive model using historical data.

12 While this is typically the case for organizations in East Africa, in other markets datasets, particularly agricultural data, may be more useful for KYC or customer segmentation purposes. Therefore, the utility of large marketwide datasets depends on the market, product and industry.

advantages: it provides rural coverage and ample historical data¹³ and it avoids the MNO channel that tends to be unreliable for many FinTech organizations (see [Box 2](#) for more detail on MNO data and the issues FinTechs experience with it). For instance, Pula leverages satellite precipitation data to estimate the likelihood of weather events. However, there is a skills gap in leveraging satellite data because it must be interpreted from the images. Moreover, another organization, Apollo Agriculture, mentioned that while machine learning has become a buzzword in the sector, it remains difficult and complex to carry out well, particularly with very large datasets like satellite data.

In sum, lack of demand for third-party data stems from a limited use case. Most organizations still use traditional underwriting processes that rely on internal data. Some organizations are exploring the use of big data to determine ability to pay in narrow use cases, such as short-term, low denomination lending. Others are testing whether big data can complement internal data to more objectively assess willingness to pay or assess willingness to pay at lower costs than current field methods. However, these methods come with additional challenges, such as validating big data sources and authenticity.

BOX 1

Credit Reference Bureaus expected to continue to play a major role in Kenya¹⁴

Nearly every lender that the research team spoke with in Kenya uses credit reference bureau (CRB) data. However very few factor it into risk—most use it to check blacklisting, double-check their own underwriting decision, or to confirm identity. Even though they use it, organizations have doubts about the accuracy and relevance of CRB data. This is because although CRBs are starting to generate credit scores based on positive and negative details provided by alternative lenders, few resource-constrained clients have obtained loans and have enough data at this time

to be included in the CRB database. It is also unclear exactly which alternative lending organizations are providing details, and to which CRBs.

Nevertheless, the increasing use of consumer data brings with it the challenge of protecting consumer data and their privacy. Consumer protectionist movements stress that digital finance providers and governments must safeguard customer privacy, whether in the form of legal frameworks and explicit customer opt-in and -out services for data usage and mining. FiDA's Snapshot 9, "[Best practices in big data analytics](#)," discusses the implications of big data on consumer protection and safety in greater depth. Moreover, the role of regulation and compliance to AML/CFT in utilizing consumer data is highlighted in an upcoming blog written by FiDA on call detail records (CDR) data.

¹³ Pula mentioned that they have access to 33 years of satellite weather data, which is available globally.

¹⁴ Very few African countries other than Kenya have accurate credit reference bureau data.

Organizations are pursuing partnerships in lieu of transactional relationships

Organizations such as MNOs that seek to monetize the data they capture about their customers are, probably for the best, treading cautiously. Not unlike many businesses globally, data-led organizations in Kenya and Tanzania are currently navigating perceived or real regulator and consent constraints. In both geographies, regulators are still determining which tack to take. In Kenya, at least at present, there is no specific data protection law, but there does seem to be a movement towards consumer protectionism. Therefore, regulation will likely be enacted in the future. Interviews revealed that some organizations in Kenya have the impression that rules on data sharing already exist—this seems to depend on whether the organization falls under a regulatory mandate. Consequently, most organizations currently obtain consent from customers as a best practice.

MNOs, particularly, expressed that they were prohibited from sharing raw data due to regulations: their interpretation was that they could not give individualized information to third parties without consent from both the individual and the regulator. This could certainly be one rationale for why they have not commercialized data to be shared “transactionally.” What appears to be emerging in the marketplace instead is partnerships between MNOs and either banks

or other FinTechs, wherein the MNO shares aggregated information (such as on a monthly basis) about the customer (once consent has been obtained), and the FinTech or bank shares a portion of the resulting revenue with the MNO.

Regulation, or the perception of regulation, is not the only driver prohibiting players from developing a transactional data marketplace. As discussed in **Box 2**, MNOs work exceedingly hard to maintain customers’ trust. In addition to violating regulations, selling raw customer data would erode this trust. Moreover, MNOs see their large customer datasets as one of their primary competitive advantages, even if they are doing very little to analyze the data at present. Without going into too much detail regarding the merits of open innovation versus impending moves by internet platform players, a partnership strategy allows MNOs to explore what is possible with data and analytics on a case by case basis, without exposing enough information to create an existential threat in the short-term. This partnership trend is likely to continue as MNOs need to rapidly build an ecosystem around digital wallets and establish partnerships with banks and FinTech players that can deliver a wider variety of digital financial services.

BOX 2

Mobile Network Operator Data—how predictive is it?

MNO data is the first bloom of big data’s potential on the African continent. Widespread adoption and use of MNO services underpin a belief that patterns in the data can reveal ability and willingness to pay, as well as demographic and other segmentation signals.

To date, MNOs have shared data, voice, and mobile

money behavior data on a case by case basis with other organizations such as FinTechs and banks. Partnerships range from loose affiliation to tightly intertwined joint ventures. And, as discussed above, most involve revenue sharing agreements that enable MNOs to monetize valuable customer data and capitalize on third-party innovation.

Interviews revealed that while a lot of faith has been placed in the value of MNO data, it has been fruitful mostly in a narrow use case of setting an initial

credit limit for a new client on a personal, short-term, low denomination loan. MNO data has limited applicability for loans of the type necessary to fund businesses or obtain workable assets.

Overwhelmingly, organizations felt that MNO data was not predictive, raising challenges even for consumer lending use cases. First, most transactions in the informal sector are still conducted using cash, even in Kenya; therefore, relying solely on mobile money data means missing significant information about an individual's expenditures. Second, as more people become aware that mobile money, data, and voice behaviors are factored into credit decisions, the data becomes distorted as people figure out ways to "game the system".

Just looking at M-PESA only shows you a small percentage of someone's total transactions. So M-PESA only gives you a small band of information. This works for MNOS who are offering very small loans, but the second you try to give out larger, meaningful sized loans you can't just rely on M-PESA data to understand someone's cash flows.

— Lendable

Finally, MNOS have shared data in a limited fashion in order to comply with regulation and, more importantly, to preserve the hard-won trust of customers. FinTechs, however, believe the motivation to limit data sharing stems more from a desire to maintain a potential competitive advantage. Several FinTechs reported that MNOS had reneged on promises or even

contracts to share data at more granular levels. Others are frustrated that more is not shared. Consequently, transactional-level details are not being "mined" by AI/ML algorithms within data analytics FinTechs, and real-time APIs exchanging data are unheard of. More commonly, data is provided at scorecard level or heavily masked; it is then factored into a decision alongside other sources of information.

MNOS don't give access to raw data, thus we rely on an analyst within the MNO to do analysis for us. There is rarely sufficient capacity/bandwidth for the analyst to carry out the level of analysis that we would like, as we are competing for bandwidth with internal and other 3rd party projects.

— BIMA

How the market will unfold in East Africa is uncertain. The forces at play—consumer protectionist movements and a drive towards open banking—must be reconciled in an environment where IT systems do not easily support APIs and the major players believe that their internal customer data is their biggest asset. Thus, in the near future, while the industry is likely to see short-term, low denomination lending products sprouting on the back of mobile operator data, it is unlikely, at least for now, that this will be a harbinger of deeper and more pervasive use cases.

FiDA will highlight how the interviewed organizations are using MNO call detail records (CDR) data and the limitations for using this dataset in greater detail in a future blog post.

In sum, organizations are treading cautiously with customer data resulting in relationships that resemble partnership structures rather than transactional marketplaces for a variety of reasons. These reasons include compliance with laws and regulations, preserving customers' trust, and maintaining an early competitive advantage.¹⁵ While the long term is uncertain—particularly as internet platform players such as Facebook and Alibaba expand market share in Africa—at least in the near term this ethos is driving slow but healthy exploration of the space.

¹⁵ FiDA's Snapshot 8, "What is the commercial landscape of digital finance?" highlights how business models and partnership structures are slowly transforming with the growing trail of big data per customer. [Snapshot 8](#) also discusses the impact that the rise of internet platform players may have on traditional digital finance providers.

Most of the organizations interviewed are still testing, refining, and experimenting with data analytics

Most of the organizations interviewed are still testing, refining, and experimenting with which datasets are most predictive and validating their analytical models. FinTechs generally reported that their algorithms relied on training data. Typically this data either does not exist for East Africa or is of low quality. This has been a driver for data analytics companies, such as Apollo Agriculture, FarmDrive, and Tala to become end-user facing (B2C),¹⁶ at least temporarily, in order to capture data that can be used to train and then directly prove the predictive power of their models. The implication of this is that they have to fund loans from their own balance sheets; as a result, the amount of capital available for lending limits the volume of data they are able to create and presents risk and loss that can be difficult for small startups to justify.

The hardest challenge is getting enough good quality data to train models. The models are extremely difficult, but that challenge is secondary to getting good, clean data....Until you reach scale and significantly prove the value of your data, you have to be B2C.

— Apollo Agriculture

The only way you can show reliability is by looking at an existing operation and performance. Bring the default rate down to a level that is allowing you to break even and then become profitable. You need a concrete example on an existing operation.

— Tiixa

Other organizations that want to remain data analytics providers have realized they have a greater chance of success if they are closer to the capturing of data. For example, First Access has developed CRM solutions that enable lenders to digitize the loan appraisal process as the first step in a broader data strategy, turning on automated analysis, alerts, and predictive analytics tools as lenders evolve.¹⁷ Additionally, Jumo has moved from being a FinTech turnkey solution to a transactional platform for partners with predictive analytics and machine learning capabilities. Even in situations in which end-user data has been shared, data analytics organizations have learned that integrating with a provider's system is painful. East African financial institutions have yet to adopt common software systems en masse, such as **Oracle Financial Services Suite** or **Salesforce CRM**. Too many different systems with which to integrate and a lack of common APIs mean that every sale requires a custom implementation. This requires tremendous time and effort and hampers the scalability of data analytics FinTechs.

Interviews also revealed that FinTechs have uncovered issues applying algorithms to the African context. As noted by Lendable, people in the United States are relatively non-correlated to each other—a bad rainy season will not affect a technology consultant in the same way that it does a farmer. However, in East Africa, because economies are still predominantly agricultural, a weather event impacts most people. Additionally, some trials with algorithms did not yield good results—either too many defaults or less predictive than traditional or current methods.

¹⁶ B2C means “Business to Consumer.” In the context of this research, “consumer” refers to the end-user of a financial service (e.g., an individual or small business receiving a loan), as opposed to another business intermediary or a bank involved in the process of lending to end-users.

¹⁷ Indeed, First Access previously had a pay-per-score model for data, but has since shifted gears to become a platform solution for lenders, offering value added services beyond a simple credit score.

We explored a few other data collaborations with third party providers for credit scoring. We found their models were not very predictive of performance for our customer base. They might work for other customers.

— M-Kopa Solar

In sum, while organizations are actively exploring how to use data and analytics, the sector is still nascent. A number of complications make it difficult for FinTechs to obtain quality data, to test analytical models, and to integrate with partners in order to leverage their data. As a result, some FinTechs have changed their business model so that they can capture data directly. Additionally, FinTechs are still refining algorithms to better fit the African context. Banks and MFIs need to improve readiness as much as FinTechs need more runway to improve models and offerings.

Most banks and FinTechs believe a business case that is strategy- and leadership-led is essential to the uptake of big data and analytics

A proliferation of data and analytics among FinTechs and other organizations should not be mistaken for progress towards financial inclusion. As mentioned above, analytics and algorithms still need refining, the large datasets available have limited proven use cases, and smaller FinTechs that are end-user facing are chipping away at an iceberg of low-income individuals and small businesses that still need financial services. It is possible that some client-facing FinTechs will reach millions—though this will require significant access to capital (for example, to finance a monthly school fee over a one month period for one million recipients would mean a total stock of approximately \$50,000,000 in outstanding loans).

Banks, however, already have access to capital, and on favorable terms. The challenge for banks is whether a strategy to reach poor and hard to reach customers is palatable. In many African countries, banks have taken the comfortable position of investing in government bonds, earning high interest rates with relatively low risk of default.¹⁸ Interest rate caps also hurt would-be low-income loan recipients; when the additional risk of default cannot be priced in, banks balance portfolios by making fewer loans to risky recipients. To make progress towards financial inclusion at greater scale, banks must be part of the solution. However, this will require a shift in mindset about who is the core customer of the bank.

Institutions and banks are skeptical of lending to rural farmers because the business model doesn't typically work. Good data on customers is not the only barrier—even if you had a perfect credit rating, the business model breaks long before it gets to data, since the basic lending cost structure of most banks and institutions leaves significant barriers to serving rural farmers.

— Apollo Agriculture

The banks and MFIs with which the research team spoke are already moving down this path. Some are undergoing a massive transformation of their core business to enable targeting mass market customers, requiring significant time and resources. All of the banks with which the research team spoke have developed a business case and expect to reach profitability from these activities. For example, banks mentioned a range of payback periods (anything from 18 months to 5 years, with ROI in the range of 25%), with no consistency beyond two organizations mentioning the same ROI number.

Indeed, one bank expects expansion to low-income segments to require a significant effort but generate low margins. Consequently, the economies of scale and scope afforded by digital are vital to the business case and digital transformation is a necessary precursor to successfully integrating data and analytics into the business model.

It's a huge transformational change—it touches every aspect of what you are doing in the business. It takes significant resources.

— FINCA Impact Finance

¹⁸ This is slightly subjective, but depending on the country sovereign bonds can be considered investment grade.

Change has to be incremental. You can't radically change business models overnight. You also can't get a massive team of loan officers to trust data overnight either.

— **First Access**

Data is essential to making the model work but organizations can't sit back and sell data and expect others to do the hard work. The things that prevent farmers from accessing credit are more than data... there are logistical and operational challenges.

— **Apollo Agriculture**

Moreover, there is a shift in mindset regarding collaboration among the banks participating in this research. They favor partnership structures because they want to develop additional products in their portfolios and they see partnership as a way to develop and launch products quickly. More importantly, the banks see FinTechs as credible partners with which to achieve these objectives. For example, Equity Bank notes that they are currently looking into partnerships with several Kenyan FinTechs—either the bank will lend on behalf of the FinTechs or, where Equity Bank is unable to take a certain level of risk, they may have a reciprocal relationship with the FinTech or financial institution.

More generally, FinTechs have felt the brunt of banks that lacked a strategy to target mass markets. Often these banks have not been prepared to use data and analytics in operations, and FinTechs have had difficulties convincing them of the value proposition and pricing. Arifu mentioned that variations in partners' readiness and ability to work with them on data analysis has been a major barrier to progress—partners often don't have the data they say they do or they lack the skill set necessary to facilitate analysis. There can also be issues in gaining trust. Like MNOS, banks have observed disruption from FinTech organizations without a clear perspective on whether a particular organization is a friend or foe.

We rarely have a client who knows what they want from the data. Often the biggest telcos and banks are a bit unsure about their strategy and what is possible.

— **FinTech organization**

In sum, to expand financial services to low-income or hard to reach customers, banks need to be part of the solution. However, before this can happen, banks need to shift their perspectives on core customers and future strategies. The highest chances of success arise from banks that focus on developing a strategy and gaining leadership support to target mass market customers first, followed by technology and process changes to deliver on the strategy cost-effectively. To this end, several organizations that were undergoing transformation revealed that so-called champions at the top of the food chain (e.g., on the board, or in leadership/management) were deciding factors in moving forward with necessary investments. FiDA's Snapshot 9, "**Best Practices in Big Data Analytics**," delves more deeply into the internal consensus and commitment that is needed—across every department in an organization, but especially at senior management level—to launch big data analytics capabilities that support digital finance.

3 Conclusions

The absence of a marketplace—and the pivoting of pioneers that were treading paths to develop the market—could indicate that it's too early for big data plays. It will likely demonstrate that a market will never emerge in the same way it has in the United States. **If data is the new oil in a digital era, most of its value—at least in Africa—will be created at the refinery, not via the process of extracting it from the ground.** Certain factors constrain a rapid expansion of activity. Progress will come, but gradually and in waves as, one by one, large players (MNOs, internet platform players, banks, etc.) open both IT and business opportunities for partnership, once the business case has been demonstrated by early adopters. Moreover, banks and MFIs can make tremendous strides by more effectively using the data they currently gather about customers.

This Focus Note has highlighted the main factors affecting the development of a market for big data in East Africa, as well as the slow but emerging demand for data analytics. First, lenders have a limited use case for third-party data because internal data has proven to be the most powerful predictor of risk. It has been used to support initial credit decisions for small denomination, short-term loans and to evaluate willingness to pay and character dimensions of the risk equation. Second, organizations are pursuing partnership models in lieu of transactional relationships. This enables them to preserve customer trust and cautiously explore new opportunities.

In the meantime, all of the banks and MFIs with which the research team spoke are either working with big data analytics providers or conducting their own analyses. They are optimistic about the eventual utility of big data; however, their recent experiences with algorithmic models suggest that FinTechs need to refine and validate their models and go-to-market strategies. Moreover, banks and MFIs need to improve their readiness for big data and analytics—in terms of systems, processes, and strategy.

Finally, data analytics cannot be detached from strategy. Interviews with banks and FinTechs suggest that leadership support to target and serve the needs of low-income and difficult to reach customers needs to come first. Only after this support can technology, such as data analytics on top of big data sources, enable the institution to cost-effectively serve this demographic. To effectively use the technology within the organization requires substantial costs and energy; without the necessary investment any efforts to serve this demographic are undermined.

While these findings draw from interviews conducted with organizations in Kenya and Tanzania, they are relevant for East Africa more broadly and potentially indicative of how the sector will evolve in other countries. MNOs are motivated to preserve customer trust across all their geographies. Additionally, it's highly likely that lenders in other African countries also rely primarily on traditional underwriting methods and internal data. Banks, MFIs, MNOs, and FinTechs in other African countries can learn from the lessons and experiences highlighted in this Focus Note to improve their readiness to adopt or deploy big data and analytics in their own contexts.

Altogether, the four factors discussed reinforce each other and are slowly progressing towards a sector that actively uses big data and analytics. They also generate new insights about the challenges organizations are facing—specifically, this is a skills and human resources issue. East Africa faces a tremendous skills shortage in terms of data scientists and experienced AI/ML professionals to do the dirty work of converting un- and semi-structured noise into meaningful signals. Several FinTechs noted both the challenges of working with algorithms and their difficulties finding and funding experts who can assist with machine learning. Even basic data analysis skills are lacking in major institutions, meaning that projects that could otherwise be implemented in a

week or a month take up to a year. FinTechs often find themselves training their partners, taking up valuable time that could be used actually delivering. Moreover, organizations that would like to move more quickly through the digital transformation process lack the expertise to see through all the systems, processes, and culture changes that need to happen.

Nonetheless, it's exciting to see some major traditional financial services organizations making strides towards digital underwriting processes that will support better financial solutions for mass market customers. It's also encouraging to hear that attitudes about FinTech players may be shifting within banks as well as MNOS. Clearly, many believe in the potential for big data and analytics to unlock value, and the level of innovation on the continent is inspiring. Organizations currently exploring or investing in big data and analytics technology will need to weigh the value of being a pioneer against the current market challenges identified. Hopefully, everyone reading this Focus Note will benefit from the candid details shared by research participants, including the internal and external data sources they use in their models and which can be found in the [Profiles of Digital Finance Organizations Leveraging Data and Analytics](#).

4 Appendix

Research approach

The Mastercard Foundation Partnership for Finance in a Digital Africa identified the following research questions to generate new information about the use of big data and analytics as well as the business results:

- How is data being used:
 - What types of data, particularly what types of big data, are being used?
 - Is this data generated internally or sourced externally from the organization?
 - What is the data used for: customer segmentation, as a factor in the risk algorithm, or something else?
- Are the benefits tangible:
 - Did the organization create a business plan to develop the capacity to work with new types of data?
 - Has the organization seen return on investment?
 - How do the results from big data compare to traditional underwriting methods?
 - What costs / risks / issues exist?
- Is a market for data developing:
 - Who is paying for data or getting paid for data directly, what organizations are involved, and what kind of data is exchanged?

To answer these questions, the research included structured interviews with 30 organizations (in-depth, in-person interviews or conference call interviews) over a 2-month period (August–September 2017), listed in Table 1. In some cases this consisted of multiple conversations with an organization to speak with individuals located in different geographies or in different parts of the business. The research team analyzed the responses to interview questions for key patterns and themes. The insights presented in this Focus Note relate to themes that were common to most of the respondents (or sub-groups within respondents, such as banks). In addition to interviews, the research team reviewed a range of relevant recent reports highlighting data and analytics issues or case examples in East Africa.¹⁹ These reports served as background to this research as well as support for preliminary hypotheses about opportunities and constraints in East Africa.

TABLE 1

Organizations participating in the research

Organization type	Research participants
FinTechs	<p><i>General</i></p> <p>Aire Labs Apollo Agriculture Lendable Arifu Craft Silicon Kopo Kopo Inc Pezesha Jumo Cignifi Tala Tiixa First Access EFL FarmDrive Umati Capital</p> <p><i>Insurance / Micro-insurance</i></p> <p>BIMA MicroEnsure ACRE Africa Pula</p> <p><i>PAYG Solar</i></p> <p>M-Kopa Solar Off Grid Electric</p>
MNOs	<p>Halotel Tanzania Safaricom Vodafone Group</p>
Banks	<p>Equity Bank Kenya Commercial Bank (KCB) Commercial Bank of Africa Tanzania (CBA)</p>
MFIs	<p>Musoni Kenya FINCA Impact Finance Juhudi Kilimo</p>

There is significant variety in the types of data being used to develop or operationalize financial services. Accordingly, the research team developed [Profiles of Digital Finance Organizations Leveraging Data and Analytics in Kenya and Tanzania](#) as part of this research. These one-page profiles stem from information gathered from organizations that agreed to share these details publicly.

A few different typologies for landscaping big data have been proposed. This research categorized data sources (big data as well as traditional) across four functional categories:

- Individuals’ financial services use or history
- Individuals’ digital interactions using a device
- Other individual data, such as psychometric survey responses
- Marketwide data, such as crop prices or satellite imagery

Table 2, below, illustrates examples of each category.

TABLE 2

Sources of data

Individuals’ financial services use	<ul style="list-style-type: none"> • Repayments data • History of financial services usage • Credit reference bureau data • Mobile money behavior • Agricultural industry transactions and receivables data (off-takers, contract farming, etc.)
Individuals’ digital interactions	<ul style="list-style-type: none"> • User activity levels, interaction, and communication trends • App downloads and usage • Social media usage • Data and voice mobile behavior • Geographic location
Other Individual data	<ul style="list-style-type: none"> • Self-reported data • Psychometric and behavioral data • Demographic data / household information • Revenue authority data • Population registration data
Marketwide data	<ul style="list-style-type: none"> • Satellite images • Agronomic practices / agricultural data • Weather, climate, and environmental records • Historical market and income data • Household surveys, financial diaries, etc.

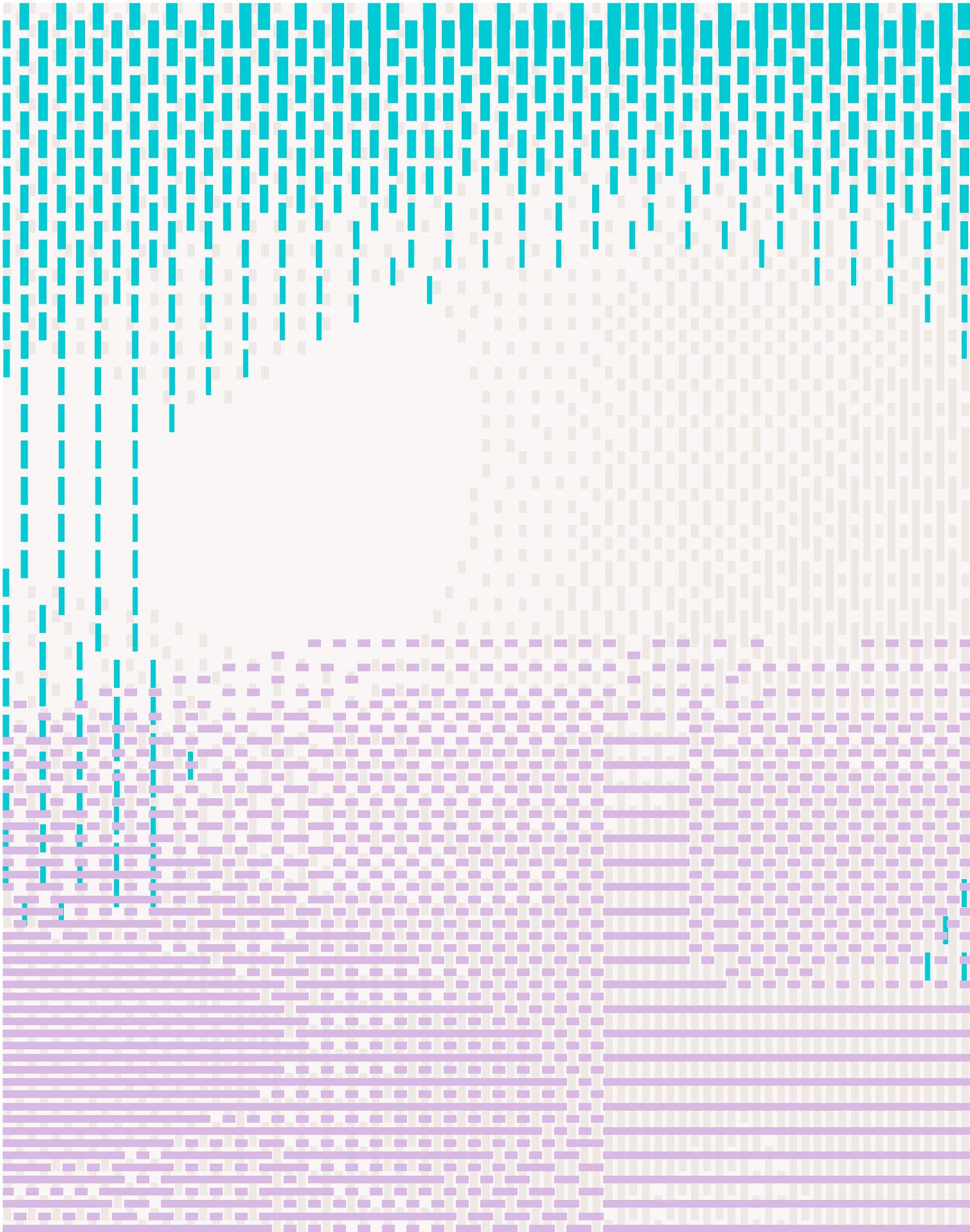
Data was also categorized by whether it was captured directly or generated internally (“internal” data) or obtained from a third party relationship or publicly available (“external” data).

The findings are limited in some important ways. The sampling focused on those organizations currently leveraging data and analytics, rather than the entire digital finance sector. This means that while the findings reflect the views of organizations with higher levels of understanding about the potential for and the reality of big data and analytics, they are not generalizable to all financial services organizations in Kenya and Tanzania. This is especially true for MNOS and banks of which, because of timing and willingness to participate in interviews, only a few were included. Nevertheless, the insights gathered reflect the perspectives of some of the more prominent financial organizations in Kenya and Tanzania. They are useful for generating findings about the current state of big data and analytics across sub-Saharan Africa and are based on candid discussions with organizations that have had the most extensive and advanced usage.

Bibliography

- Accion Global Advisory Solutions. 2017. "Unlocking the Promise of Big Data to Promote Financial Inclusion." <https://www.accion.org/content/unlocking-promise-big-data-promote-financial-inclusion>.
- Accion Venture Lab. 2016. "Bridging the Small Business Credit Gap Through Innovative Lending." Accion Venture Lab. <https://www.accion.org/content/bridging-small-business-credit-gap-through-innovative-lending>.
- Arabehty, Pablo Garcia, and Jacobo Menajovsky. 2016. "Those Star-Ratings Do Matter for Financial Inclusion." CGAP. September 23, 2016. <http://www.cgap.org/blog/those-star-ratings-do-matter-financial-inclusion>.
- Bloomberg New Energy Finance, and Lightning Global. 2016. "Off-Grid Solar Market Trends Report 2016." World Bank Group.
- Caire, Dean. 2017. "Digital Financial Services for Smallholder Farmers: What Data Can Financial Institutions Bank On?" Nairobi, Kenya: Mercy Corps Agrifin Accelerate. http://mercycorpsafa.org/wp-content/uploads/2018/01/Digital-Agro-Data-Edited_Final.pdf.
- CGAP. 2016. "An Introductory Course to Digital Credit." CGAP. <http://www.cgap.org/blog/introductory-course-digital-credit>.
- Chen, Greg, and Xavier Faz. 2014. "Hype or Hope? Implications of Big Data for Financial Inclusion." CGAP. September 4, 2014. <http://www.cgap.org/blog/hype-or-hope-implications-big-data-financial-inclusion>.
- . 2015. "The Potential of Digital Data: How Far Can It Advance Financial Inclusion?" Focus Note No. 100. CGAP. <http://www.cgap.org/sites/default/files/Focus-Note-The-Potential-of-Digital-Data-Jan-2015.pdf>.
- Cook, Tamara, and Claudia McKay. 2015. "How M-Shwari Works: The Story so Far." CGAP. http://www.cgap.org/sites/default/files/publications/multimedia/CGAP_Annual_Report_2015/files/forum/forum-how-m-shwari-works-apr-2015.pdf.
- Costa, Arjuna, Anamitra Deb, and Michael Kubzansky. 2015. "Big Data, Small Credit: The Digital Revolution and Its Impact on Emerging Market Consumers." Omidyar Network. https://www.omidyar.com/sites/default/files/file_archive/insights/Big%20Data,%20Small%20Credit%20Report%202015/BDSC_Digital%20Final_RV.pdf.
- FIBR. 2016. "Briefing Note – The Lean Product Waltz: Data-Driven User Research & Research-Driven Data Analysis." FIBR. <https://static1.squarespace.com/static/5682a94969a91ad70f60a07c/t/592e5cofd2b857da93629f11/1496210450638/FIBR+briefing+note+2-Lean+Product+Waltz+FINAL+12262016.pdf>.
- . 2017a. "Briefing Note on Lendable: Case Study of a Marketplace Lending Platform in East Africa." FIBR.
- . 2017b. "Briefing Note – Alternative Lending: Landscaping the Funding Models for Lending FinTech Companies." FIBR. <https://static1.squarespace.com/static/5682a94969a91ad70f60a07c/t/596435684f14bcbab01c5069/1499739497868/FIBR+Briefing+Note+3+-+Alternative+Lending+July+2017.pdf>.
- GSMA. 2014. "Zoono." Mobile for Development. November 3, 2014. <https://www.gsma.com/mobilefordevelopment/programme/mobile-money/zoono>.
- Herrera, Miguel. 2014. "Leveraging Mobile Phone Data: Tiixa's Balance Advance." CGAP. September 15, 2014. <http://www.cgap.org/blog/leveraging-mobile-phone-data-tiixa%E2%80%99s-balance-advance>.

- Kaffenberger, Michelle, and Rafe Mazer. 2014. "Simple Messages Help Consumers Understand Big Data." CGAP. September 29, 2014. <http://www.cgap.org/blog/simple-messages-help-consumers-understand-big-data>.
- Mercy Corps. 2016. "Fertile Ground Series: A 2016 Tanzania Ecosystem Review & Strategic Perspective on Digital Financial Services for Smallholder Farmers." Mercy Corps. <http://mercycorpsafa.org/wp-content/uploads/2016/12/Tanzania-White-Paper-2016.pdf>.
- M-KOPA Solar. 2016. "Digital Insights Build Trust and Enable Growth." M-KOPA Solar. <http://www.m-kopa.com/digital-insights-build-trust-enable-growth/>.
- Muench, Dirk, Daniel Waldron, and Xavier Faz. 2016. "Access to Energy and Finance: An Integrated Approach." CGAP. <http://www.cgap.org/sites/default/files/Working-Paper-Access-to-Energy-and-Finance-April-2016.compressed.pdf>.
- Naef, Ed, Philipp Muelbert, Syed Raza, Raquel Frederick, Jake Kendall, and Nirant Gupta. 2014. "Mobile Data for Development." Cartesian. https://www.cartesian.com/wp-content/uploads/2015/07/Cartesian-Mobile-Data-for-Development_July2014.pdf.
- Owens, John, and Lisa Wilhelm. 2017. "Alternative Data Transforming SME Finance." IFC.
- Ralph, Oliver. 2017. "Insurance and the Big Data Technology Revolution." Financial Times. February 24, 2017. <https://www.ft.com/content/bb9f1ce8-f84b-11e6-bd4e-68d53499ed71>.
- Rizzi, Alexandra, Isabelle Barrès, and Elisabeth Rhyne. 2017. "Tiny Loans, Big Questions: Client Protection in Mobile Consumer Credit." The SMART Campaign . http://smartcampaign.org/storage/documents/Smart_Brief_Tiny_Loans_Big_Questions_Final_092617.pdf.



Partnership for **FINANCE**
in a **DIGITAL AFRICA**