

# Now you see me:

How alternative data is unlocking new markets for financial services

Phone address book includes first and last names

Pays DSTV bill on time every month

345 Facebook friends

Overnight airtime balance: 232 Ksh



Advancing Financial Inclusion

The focus of this note is on the increasing adoption by the private sector of alternative data sources and analytical methodologies for financial service provision. It highlights the major trends that are contributing to this increasing adoption and their relevance for financial inclusion.

Further it identifies different categories of financial service provision based on the degree to which new data sources and methodologies have been adopted and the barriers and drivers to further adoption. In doing so, it aims to identify what is possible for different categories of providers with regards to financial services provision based on new data and analytical trends.

The note draws from research conducted by insight2impact (i2i) for a scoping survey on the use of client data and analytics by financial service providers (FSPs) and other related organisations. It included desktop research on existing studies, new opportunities and applications, as well as in-person and telephonic interviews with 46 FSPs, fintech start-ups, analytical providers, consultancy firms, design organisations, and academic or research organisations. For a full list of stakeholders included in the study, see the Appendix.

The results of this survey are a starting point for how new and existing clients can be better served through these new data sources and analytical methodologies and how these developments can impact financial inclusion.

# Contents

Trends in data and analytics	3
Impact on the provision of financial services	4
Emerging categories of financial service provision and inclusion	4
Drivers and barriers to adoption	9
Going forward	11
Appendix: Organisations	13
Bibliography	14

## Authors:

Christiaan Loots  
Krista Nordin  
George Mason  
Abenaa Addai  
Nkosi Ncube

# Trends in data and analytics

Financial service providers are increasingly adopting historically non-financial data (“alternative data”) as a way to analyse current and predict future financial behaviour to inform the design and distribution of financial services. There are two major trends that are giving rise to this:

**The increase in the number of sources and volume of alternative data available.** Digital data is growing and it’s doing so quickly. Every day 2.5 quintillion bytes of data are created. In the past two years alone, 90% of the data in the world today was created (IBM, 2016). This is set to continue and the digital data in the world is predicted to increase by 4400% between 2009 and 2020. This is likely to have a big impact in the developing world, specifically in Sub-Saharan Africa where the volume of data generated by mobile phones alone is set to increase by more than 200% between 2016 and 2019 (GSMA, 2015).

The following factors are key drivers of this increase:

- **Mobile and online footprints of customers are expanding rapidly.** The number of unique mobile subscribers in Sub-Saharan Africa is set to increase from 386 million to 518 million between 2015 and 2020. A big impact on the volume of available data will be observed in the migration to smartphones. Smartphone ownership is set to increase from 160 million to 540 million over the same period with 55% of all mobile connections in Sub-Saharan Africa to be made through these devices by 2020 (GSMA, 2015). This continues to enable more and more adults to become broadband users. In 2014, the number of monthly active Facebook users in Africa exceeded 100 million (Constine, 2014). In 2015 the number of internet users in Africa was 331 million, or 28.6% of the population (Internet World Stats, 2015). The decreasing cost of smartphones will continue to drive these trends and increase the amount of available data.
- **Rise of the Internet of Things (IoT)** that embeds electronics, software, sensors, and network connectivity into everyday objects to enable them to collect and exchange data, will add to the amount of data that is available. By some estimates, the number of connected devices worldwide which capture sensor data will reach 30 billion by 2020 (IDC, 2014). While most of this is still happening in more developed countries, there are pockets of innovation in the developing world. For example, M-KOPA, a Kenyan solar energy company selling home solar systems in Kenya, Tanzania, and Uganda, generates data from sensors in their solar panels, which is then used in their loan decision making.
- **The open data movement** led by civil society organisations and governments is making public databases and government data more accessible. In 2013, 55% of countries in a 77 country sample “have developed

some form of Open Government Data initiative” (Davis, 2013). Further, call data from more people using mobile phones, mobile money transactional data from more people having accounts and satellite generated imaging through google earth are a few examples of data exhaust – data generated as a byproduct of people’s digital activities – and open data initiatives which also increase the amount of available data.

**Analytical capabilities are increasing to more accurately predict future financial behaviour based on alternative data.** Processing power, software, and analytical methodologies are advancing at an exponential rate. Moore’s Law observes that over the last 50 years, processing power has doubled roughly every two years. These advancements are allowing providers to more accurately predict future financial behavior, based on alternative data, than they could in the past. The following drivers are contributing to this:

- **Smarter software** offers new ways to utilise and analyse larger and varied datasets that will make it easier for non-specialised analysts to use.
- **Cloud connectivity** and the ability to **access analytical software** on a pay-for-use basis lowers cost, allows for mobility, ad hoc scalability and faster deployment. This gives smaller businesses or start-ups access to software and services that were previously inaccessible (C3IT, 2016).
- Advances in analytical capability are opening up **new fields of analysis** previously not widely applied within the business environment. For instance, machine learning is increasingly being used to develop complex models and algorithms based on historical relationships and trends for predictive analytics. This allows analysts to discover insights in a more accurate, reliable, and repeatable way.
- The capability to analyse **unstructured data** like text, images, audio and video to extract insights is becoming more accessible. Examples include natural language processing, text mining and image or video analytics. These advances allow for additional layers of information which can be used to analyse and predict future financial behaviour.

The increase in the volume of alternative data and capability of financial service providers to use it to predict financial behaviour is likely to continue, making alternative data sources and analytical methodologies more available and less costly.

# Impact on the provision of financial services

The availability of more alternative data and advanced analytical capability to use it, has the potential to create new markets for financial services, improve the consumer experience when using financial services and allow providers to enter these new markets viably.

**Creating new markets for financial service provision.** In developing countries, the lack of traditional financial data on the majority of the population has contributed to poor financial service provision and high levels of financial exclusion. In the absence of this data FSPs, especially those offering credit and insurance, are unable to properly understand their customers and assess their risk, either forcing them to charge high interest rates to protect against unforeseen risk or discouraging them from serving new markets. Most formal traditional credit providers rely on credit bureau and transactional data to verify the identity of applicants and assess their ability and willingness to repay a loan. However, credit bureaus only have data on a very small portion of the adult population: 93% of adults across Sub-Saharan Africa do not have credit bureau data (World Bank, 2015). The result is that the vast majority of the adult population is excluded from access to formal credit. The proliferation of new, alternative data sources and improved analytical capabilities has the potential to change this paradigm by overcoming information barriers (and associated risks) and creating new markets for financial service provision.

**Improving the consumer experience when using financial services.** New data sources and innovative analytics enable FSPs to gain a more in-depth understanding of each customer. These new insights can be used to improve their consumer experience by offering services that are more flexible, and customised to the needs of the individual. Customer data and analytics also create additional opportunities for FSPs to deepen and enhance customer relationships and provide more value through better products, customer experiences and value added services. This deeper understanding of customers also lowers the costs to providers associated with identity verification and the on-boarding of new customers. These opportunities with alternative data are also observed to some degree in payments and savings providers. While reducing information asymmetries is less of an issue for products like savings and payments, data and analytics are being used to improve FSPs' understanding of customer behaviour and to inform agent management and distribution through a better understanding of demand. For all product categories, data and analytical methodologies are also being used to improve existing business processes which are not directly related to customer experience (for instance fraud management).

**Offering scalable solutions for financial services provision.** Most scalable solutions have to incorporate some degree of automation in the decision-making process. For years this limited the types of data that could be used in the decision-making process. Organisations generated large amounts of unstructured data (like reports, product descriptions, emails, notes etc.), which could not typically be used in automated decision making as the information was not captured in a systematic way and the analytical tools to analyse it did not exist or were not commonly available. Therefore automated business decision-making processes used structured data as a primary input, while unstructured data was largely limited to being used qualitatively to enhance automated decisions.

Recent trends in the availability of more data and powerful analytical methodologies and software has resulted in the improved accuracy of automated decision-making processes by adding more of the same information (more of the same data enhances statistical accuracy), and/or by adding new types of information – all of which can either be generated by the provider themselves, or can be obtained through external (new) sources. These new types of information can either be related directly to the provision and use of financial services, or it can be completely unrelated. The latter holds a lot of potential value for financial inclusion, as it provides a way for providers to overcome the information barriers which previously have led to the exclusion of large portions of the population (“information invisible” customers).

## Emerging categories of financial service provision and inclusion

The use of alternative data in business decision making has grown considerably over the last few years, specifically for credit products. At the end of 2015, TransUnion conducted a benchmark study on the state of alternative data that included more than 300 lenders in the USA and found that just over a third (34%) of all lenders were using alternative data in their credit decision process. Given the above trends, the adoption of alternative data will continue to increase and become more widespread across the industry and different product markets.

This section looks at the degree to which alternative data is used by providers to date. It identifies four categories of usage based on the type of data and analytical methodologies used. For more information on how the categories were identified, see Box 1.

## Box 1: Terminology used (What's new?)

**For the purposes of this note, traditional versus new is defined as follows:**

**Traditional data types** refers to structured financial data generated internally by financial service providers (transactional, repayment, claims data, etc.) or externally by traditional industry data providers (like credit bureaus) that is used for analytics which inform financial services decision making. It also includes data generated from direct engagement with clients (surveys, interviews, etc.) and personal client data (income, demographic, etc.). Call data records, as well as transactional data or repayment data captured by mobile network operators or mobile money providers would also fall in this category.

**New data types** refers to structured, semi-structured or unstructured data not traditionally used for analytics which inform financial services provision decision making, and not necessarily related to past financial services usage. Examples include social media data, psychometric data, satellite imagery, geospatial data, mobile application data, etc. Although these are not directly related to financial services use, they are being used in new analytical processes to inform financial services decision making. These new data types provide insight into qualitative concepts like character, behaviour, and social connectedness, allowing providers a broader understanding of individual customers.

**Traditional data usage** refers to the application of long-established analytical methodologies or processes for conventional business purposes. For instance, the use of traditional data sources in the development of a traditional credit scoring model using non-iterative, standard statistical methodologies (like regression analysis) is considered traditional usage for the purposes of this note.

**Innovative data usage** includes organisations that obtained more or new data in a new or innovative way (for instance through partnerships). It also refers to the use of advanced analytical methodologies or processes (like machine learning, artificial intelligence, natural language processing, etc.) which allows for the use of more data or new data types.

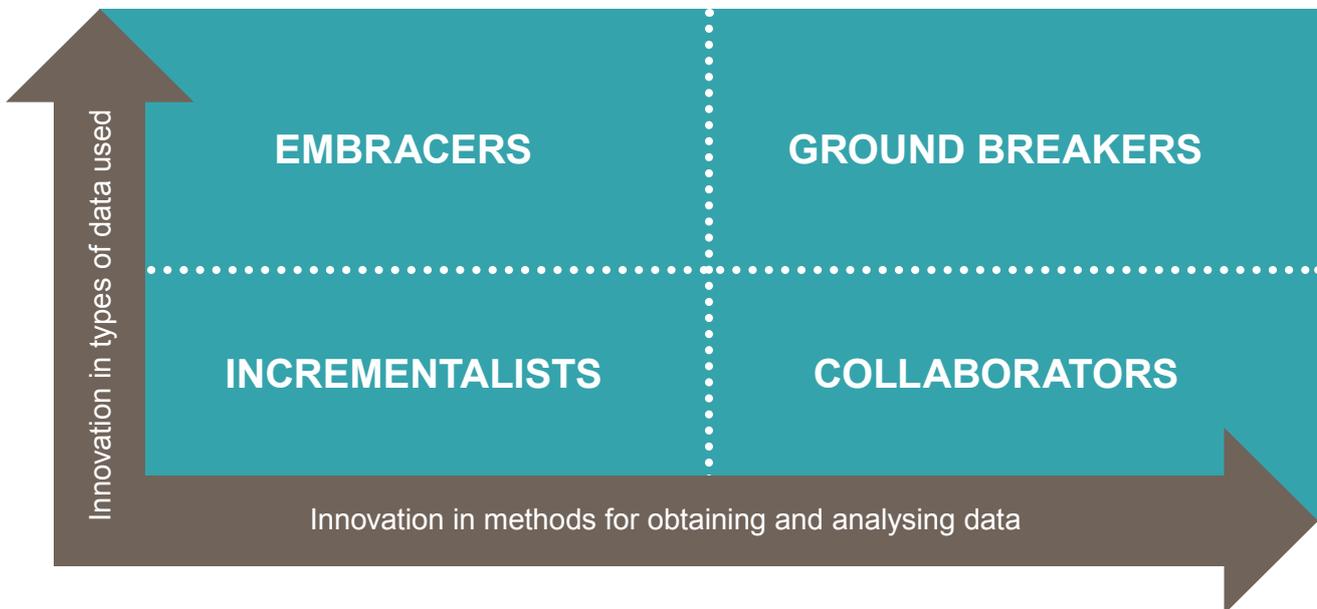
**Source: Author's own**

Financial service providers are using alternative data to varying degrees. Figure 1 below plots the organisations reviewed in this study by the degree to which they are using more or new types of data in financial services decision making and the degree to which this data has been obtained or used in a new or innovative way. The latter includes:

- the use of new analytical methodologies to utilise new data;
- new partnerships to obtain new data; or
- the choice of delivery models in order to utilise the potential of new data or new analytics.

---

<sup>1</sup>TransUnion defines alternative data as any information that is not captured in a traditional credit score or data points that are incremental to the credit bureau report. Adoption has been lower in savings, payments and insurance.



**Figure 1: Four categories of organisation, based on the nature of their alternative data usage**

Source: Author's Own

**1. Incrementalists** are located in the bottom left of the spectrum and use traditional data sources with long-established methodologies for conventional business decision-making processes. For example, using regression analysis to segment customers based on the providers' transactional data, customer data captured by the organisation and industry shared data such as credit bureau data.

Large companies in this group may be experimenting with new sources or methodologies but their core business is still reliant on traditional data and methodologies. The use of new data

sources or analytical processes is more incremental in nature and is applied to improving existing processes or set up as experimental units.

Other providers are acting as incubators or accelerators for innovative start-ups to create and connect fintech oriented innovation ecosystems and to bring scale to new ideas. For example, Barclays in Africa is funding the Barclays Rise Fintech accelerator in South Africa. More information can be found in Box 2 below.

### Box 2: Barclays Rise FinTech Innovation Hub

Rise, created by Barclays, is a physical and digital community that provides fintech start-ups and businesses with the connections and resources to create ground-breaking businesses. It is about connecting the world's most brilliant innovators and start-ups to each other, to corporates and to resources. Rise has innovation hubs in London, Manchester, New York, Cape Town, Tel Aviv, Mumbai and Vilnius. The Barclays Accelerator, powered by Techstars, is run from these hubs. Barclays Accelerator is an intensive 13-week programme designed to accelerate promising fintech start-ups.

For more information on the Barclays Rise FinTech Innovation Hub, please visit: [www.thinkrise.com](http://www.thinkrise.com)

Source: Author's own

These providers may also be thinking about improving existing or developing new data strategies, data management and data governance, as well as reacting to the requirements that the big data revolution is imposing on data infrastructure and system requirements. In general though, these players are slower to adopt these innovations as their business models are set, business processes take a longer time to execute and they have sunk costs associated with legacy systems, which restrict innovation.

**2. Collaborators** are located in the bottom right of the spectrum. Collaborators are traditional players who access new data by partnering with other organisations, typically not in the same sector, to gain additional insight into their customers which they can then use in the design of new services or to reach new markets. Collaborators have some of financial inclusion's biggest success stories. The most common of which have been between traditional banks and Mobile Network Operators (MNOs). MNOs have large digital databases on their customers, and through partnership allow banks insight into segments of the market that are not their direct clients (as well as additional information about their existing clients). For example:

- In **Kenya**, the **Commercial Bank of Africa (CBA)** and Safaricom partnered to offer credit services (M-Shwari) to M-Pesa's 20 million users. M-Shwari uses historical data from Safaricom on mobile phone and M-Pesa usage to assess creditworthiness and offer loans to its customers. This historical data, including airtime purchases, airtime credit, M-Pesa balance, M-Shwari savings activities, call records, data usage and the customer lifetime length, is used to construct a credit score and credit limit. To date, CBA have disbursed more than 20 million loans to almost 3 million unique borrowers through M-Shwari. CBA reports a 90 day non-performing rate of only 2% on M-Shwari loans (Cook & Mckay, 2015).
- In **Zimbabwe**, **EcoNet Wireless** went as far as to buy Steward Bank to give them a license to offer credit through their EcoCash Loans product. Similar to M-Shwari, EcoCash uses transactional data from the EcoCash wallet and EcoCash Save Account, along with information on airtime and data purchases to calculate an applicant's credit limit. More information on this example can be found in Box 3 below.

### Box 3: EcoCash loans example

Econet Wireless is an MNO in Zimbabwe that had 10 million subscribers as of the end of 2015 (Musarirwa, 2016). Launched in 2011, EcoCash is a subsidiary of Econet Wireless that offers person to person (P2P) payments to 5.8 million of Econet's subscribers (Crowley, 2015). Since its launch, EcoCash has become a fully-fledged financial services provider that offers payments, credit, savings and insurance products in partnership with Steward Bank, another wholly owned subsidiary of Econet Wireless, which they purchased as a deliberate move rather than partnering with a third party entity.

Steward Bank hosts the savings and credit accounts offered as EcoCash Save and EcoCash Loans respectively, but the accounts are accessed by customers entirely on their mobile device and do not require any branch engagement. The EcoCash Loans product is offered exclusively to EcoCash customers who also saved at least USD 5 monthly over three months. The EcoCash Loan model does not in any way rely on traditional collateral, payslips or credit information from a credit bureau to determine credit risk. Instead, they use several innovative data sources, including transactional flows (value and volume) through the EcoCash wallet, transactional flows through the EcoCash Save Account as well as monthly expenditure on airtime and data. Furthermore, no identity verification is required as EcoCash maintains a record of the SIM card registration of all its customers, which is sufficient for the loan application.

**Source: Author's own**

<sup>3</sup>M-Pesa is a mobile payments product offered by Safaricom, a leading mobile network operator in the country. M-Shwari has 4.5 million users, and can only be accessed through the M-Pesa mobile platform. <sup>4</sup>Please note, although EcoCash Loans relies on an innovative business model to access innovative data sources, the nature of the data they use is similar to traditional data types - for instance transactional data or repayment data. For the definition of traditional data, please see Box 1.

**3. Embracers**, located in the top left of the spectrum, recognise the need and power of incorporating new approaches into traditional business practices. They have either adopted new sources of data or new analytical methodologies. The former combine new sources of data with traditional sources of data to add additional layers of information to analytics to improve their financial service offerings. Some of these organisations provide financial products directly to customers, while others provide services to FSPs.

They are able to supplement traditional data used in large FSPs with new and emerging data sources. For example:

- In the **USA**, **Beyond the Arc** uses data on consumer complaints made public by the regulator as well as social media data to enhance FSPs' transactional and customer relations management databases.
- In **Kenya**, **Superfluid Labs** enhances their client's existing customer profiles by combining internal client data with third party data (e.g. credit bureau and MNO data) and public data sets such as social media.

The latter recognise the need for and power of using innovative analytics, such as machine learning, as a means for aggregating more traditional data sources and updating large customer databases more efficiently and at a lower cost. For example:

- In **South Africa**, **E-Intelligence** uses a combination of customer data from their FSP clients and traditional credit bureau data to create and maintain large customer databases which are kept up to date through advanced analytics which identify only variables and data points that are most likely outdated. This keeps the cost of aggregation low.

**4. Ground breakers** are located in the top right corner of the spectrum. They use both new data sources and analytical methodologies. Many of these organisations are focused on creating new markets and tend to be in the fintech start-up or entrepreneurial space. Over the last few years, there has been a proliferation of these dynamic organisations capitalising on the opportunity of both new data sources and innovative analytics to rethink traditional models of financial service provision. These organisations span across the entire value chain of data and analytics. There are four sub-categories of ground breakers identified in this study.

The first sub-category features organisations that provide financial services directly to customers by analysing new data types with cutting-edge analytical methodologies. For example:

- **Jumo**, an African mobile money marketplace provider, uses over 10 000 variables derived from their mobile network and mobile telephone use data, as well as machine learning

methods to assess a customer's credit risk. Data points used include calling records (how many and to whom), airtime usage (amount and top up location), the type of phone used, whether text messages are being sent, data purchases and mobile money transactions.

- **Branch**, a digital credit provider, scrapes an applicant's mobile phone for thousands of data points, including their contact list, phone make and model, and GPS location when a customer downloads Branch's mobile application and applies for a loan. Through the use of machine learning algorithms, they then assess creditworthiness and have the applicant's loan in their M-Pesa account in minutes.
- **Segovia**, an innovative bulk payment provider, uses satellite imagery and biometrics, coupled with imaging analytics and machine learning to better deliver payments to the recipients of development or aid programs.

The remaining sub-categories use both new data sources and advanced analytical methodologies, but do not provide financial services directly to customers. Rather, they sell business services to FSPs.

The second sub-category is a group of emerging alternative credit decision service providers who are analysing alternative data sources such as psychometric data, social media data, satellite data and mobile phone data, using machine learning methods, in order to assess an applicant's creditworthiness.

- **Entrepreneurial Finance Lab (EFL)** is a leader in using psychometric credit scoring to enable better lending decisions. According to EFL, "Psychometrics refers to the measurement of knowledge, abilities, attitudes and personality traits". They collect indicators on these measures from the way that an applicant fills in the application form (rather than just the data from the application form itself). Using advanced analytical modelling (including machine learning methods), they then use psychometric characteristics to predict the character, willingness, and ability of applicants to repay loans (characteristics that are predictive of credit risk). This new dimension to credit risk scoring allows them to assess and manage an individual's default risk, regardless of whether they have a credit bureau record or collateral (EFL, 2016).
- **Social Lender** is an up-and-coming lending solution based on an applicant's social reputation on social media platforms. Currently operational in Nigeria in partnership with the Sterling Bank, Social Lender develops social reputation scores using social media data and its proprietary algorithm. Loans are guaranteed by the user's social profile and network, allowing users to then borrow funds based on their social reputation rather than a traditional credit score. Users can access their money via existing banking channels and the company is exploring how to disburse credits through mobile money.

<sup>5</sup>Segovia's platform integrates seamlessly across local payment provider networks and allows their partners to schedule, issue, reconcile and track every payment. They also automate routine tasks, and provide real-time analysis on their disbursements and recipients.

Social Lender has over 10 000 users, and having recently graduated from the Barclays Rise Accelerator program, Social Lender is patent-pending and gearing up to expand service beyond Sterling Bank and outside of Nigeria.

- **Lenddo** is an innovative analytics company using new data types and machine learning technology for social identity verification and credit scoring. Lenddo has developed a business model that relies on the applicants social networks (such as Facebook, Twitter and LinkedIn) to predict credit worthiness. The information obtained can include network data, messages, tagged photos, browsing data and search data. Lenddo uses a variety of analytical methodologies including machine learning methods to construct algorithms and analyse data. The Lenddo credit score assesses the applicant's willingness to pay on a scale of 1-1000, with a minimum acceptable score of 300. The applicant must have three trusted connections within their network with the minimum acceptable score. The credit score of each of the connections is directly linked to that of the applicant, meaning that a default or delinquent payment will have a negative impact on the scores of the connections as well as the applicant. Using its proprietary algorithm, Lenddo is producing results on par with, or more accurate than traditional methods.
- **InVenture**, whose mobile application scans applicants' phones for detailed data points, such as the proportion of their contact list with both first and last names, to predict credit risk and provide credit decisions within 20 seconds. Interestingly, InVenture has found that if at least 40% of an applicant's contact list were organised with both first and last names, they were 16 times more likely to pay on time.

The third sub-category is a group of analytics providers – for example **Brighterion** and **Demyst data** – where the entire business is designed around the aggregation and selling of new sources of data and insights. These organisations use innovative analytics, which include machine learning, and face few limits with respect to what they can do with data; they can integrate and analyse any data source whatever the format.

The fourth sub-category contains research labs – for example **USC Data Science Lab** and **iHub Data Science Lab** – that provide analytics services and actionable insights to their clients by applying machine learning techniques to a range of new data sources.

## Drivers and barriers to adoption

If new data sources and analytical methodologies offer ways to overcome traditional barriers to financial inclusion and are being adopted by some market players, the question remains why adoption is not happening faster or on a broader basis. Why

are new opportunities not being more disruptive? This section explores the drivers and barriers of adoption.

- **Adoption drivers:** When an innovation becomes available (rolled out in the market), it is adopted by other organisations because market players are **aware of it, understand the possibilities and potential value** that it allows and are willing to take the risk of experimenting with it (or can convince others that it is worth the risk).
- **Adoption barriers:** There are a number of barriers that can impede this adoption. For example, inertia or **short-term economic constraints** or the way that individuals and organisations discover, digest, accept, **respond to and act upon new information**<sup>6</sup>. Mostly though, for new data sources and analytical methodologies, barriers are around implementation. This includes lack of defined business strategies, business questions, data leadership, data sourcing and integration, scalability and regulation around the collection, storing and use of data.

Often these barriers and drivers to adoption are two-sides to the same coin. For example the lack of a business case can be a barrier, while the confirmation of a business case be a driver. The following considers both sides of these coins for the key issues for the adoption of alternative data.

**Awareness of possibilities increasing, understanding of importance lagging.** Awareness of new possibilities that are available because of trends in client data and analytics is increasing. This is true for traditional FSPs and non-traditional providers.

- For **traditional FSPs**, particularly bigger ones, senior managers are starting to understand the value of data and analytics in decision making.
- For **non-traditional providers**, fintech firms and start-ups are embracing new possibilities, data sources and analytics are shaping entirely new models for the delivery of financial services.

However, there are still organisations and managers that do not yet value the importance of data in decision making. They are either not aware of or do not see the need for a data strategy or data leadership. Others are overwhelmed by the sheer number of innovation opportunities and do not know where to begin. This is particularly the case for innovations like the use of social media data which are not yet fully understood.

**Uncertainty around which models will prove most viable.** In a recent survey of credit providers based in the USA (TransUnion & Versta Research, 2015), evidence on the outcomes of the use of alternative data was identified as a major factor which would convince non-users to adopt alternative data. This included

<sup>6</sup> This process can be broken down into three broad categories: Information processing barriers, behavioural barriers, and implementation barriers.

evidence of risk assessment improvements (66% of non-users), evidence on the creation of new profitable markets (62%) and peer group case studies, which include cost benefit analysis on using alternative data. Due to the recent nature of innovation in this space, organisations are still experimenting with new data sources and methodologies, and it is still unclear which options and models will prove viable or most successful. While a number of use cases have been identified, it will take time to confirm which options are most feasible, add the most value or are most successful. However, there are some promising signs. For example:

- Most start-ups providing credit based on alternative data sources and innovative analytics are still in the trial phase. However, (Costa, et al., 2015), reports that “several firms have suggested that their algorithms now outperform traditional credit scoring algorithms (such as FICO scores) in the markets they target, even though FICO and similar industry incumbent scores have had decades of refinement”.

**Sunk costs and short-terms economic constraints restrict ability to invest.** The use of most data sources and analytical methodologies require investment in processes, systems, software and hardware. Most large FSPs face sunk costs in existing legacy systems and are focused on short-term goals or limited by short-term economic constraints. This results in limited ability or willingness to invest in new options, even in cases where the long term benefits may have been proven already. For example:

- **Lenddo** reported that they incurred significant costs in the initial data gathering, development and training of their algorithm and the initiative took several years.

**Shortage of required skillsets drive up the cost.** The complexity of applying new analytical methodologies and integrating data sources requires a unique combination of skill sets, including an in-depth understanding of mathematics and statistics, as well as an understanding of information technology, data, business models and business practices. This combination – known as “data science” – is gaining prominence very fast. However, the industry is still catching up in procuring or developing the required skill sets and as a result there are fewer staff with these skill sets available at a higher cost. The availability of off-the-shelf and smarter software that allows for analysis by non-specialised analysts will help to ease the effects of the skills constraints in this space. Ultimately though, the interpretation of insights for business purposes still requires a human eye combined with some business experience.

**Most organisations lack defined data-driven business strategies.** A lot of hype has been created around the possibilities that the trends in data and data science unlock, and organisations that are exploring this space face pressure to remain competitive in this environment. Organisations are therefore focused on collecting lots of data and applying complicated analytical methodologies, without necessarily identifying what they are trying to solve. This results in insights that are not specific to business needs and organisations that do not know what to do

with the insights generated; resources are being used inefficiently in the process. Successful initiatives start with identifying specific business questions or needs, identifying data requirements and availability (by either studying existing/available data, or gathering new data related to data requirements for the specific need) and then choosing a suitable analytical methodology.

**Data is becoming more available, but potentially not more accessible.** Data is becoming available at unprecedented levels but some companies that own and generate data (especially transactional data) are protective of their data as they see it as a business asset. The increasing perception of the value of data results in many companies building walls around their data to prevent third parties from accessing it or the overvaluing of data when partnerships are considered. This makes it more difficult for companies who rely on data owned by third parties to get access to suitable data. Countries like the United Kingdom are therefore considering how banks can open up access to their customer data to other businesses to encourage innovation. For example:

- **The UK Treasury** “tasked an industry-led Open Banking Working Group (OBWG) with developing a new framework for underpinning an open banking standard”, which could be fully operational by 2019 (Out-law.com. 2016).

In an environment where it is becoming easier to build one's own database from scratch through new channels like mobile applications, companies that are protective of their data run the risk of missing business opportunities through partnerships. Further open data initiatives driven by governments or civil society organisations are “not always accessible, easy to use, machine readable, up-to-date or free of cost” (Far & Chan, 2016). This could undermine the potential value and use of this data.

**Need for executive level champions for data initiatives.** The increasing volume, velocity and variety of data generated today are placing constraints on the management and use of both existing and new data sources. The adoption of alternative data sources and advanced analytical methodologies is made more difficult in larger organisations by the fact that data is often generated, managed and analysed in silos, making an overall enterprise view of the data (and customers) difficult or impossible and effectively too costly. Organisations are therefore placing more emphasis on data as a cohesive focus area within the organisation and undergoing structural changes that reflect the need for leadership on data including the creation of the executive (or C-suite) positions such as the Chief Data Officer.

**Sourcing and aggregating data is resource intensive.** Access to relevant data is only a starting point. Data integration and data quality is critical for any data-driven strategy. While this is relevant for new sources of data and for existing data, the increasing variety of alternative data sources means that data integration and quality will become a bigger factor in the success of new data and analytics initiatives. To some degree, ongoing data integration from known sources can be automated, but it still requires time, resources and human interpretation to add a new data source to your dataset as well as to maintain

existing integration processes as variable names etc. may change. Organisations such as **Demyst Data**, **Enigma** and **FirstAccess** identified data integration and quality as a field that will increasingly require skills and human input, even more so than for the application of new analytical methodologies. Best practices and standards in terms of data formats, naming conventions and data sharing could therefore facilitate easier sharing for users relying on multiple external data sources.

**Varying levels of scalability of models.** While most new models utilising alternative data or innovative analytics are still in the experimental phase, some have proven to be more scalable than others. For example organisations like **Lenddo** – which incurred significant costs and time to develop their model – have found that their algorithm is applicable regardless of the country context and can therefore be easily implemented in other countries. Further the marginal cost of replicating their algorithm in a different context (or country) is low.

However, to develop an accurate algorithm is expensive as noted above. First you need to train it using lots of data, covering both good and bad outcomes. If you do not have existing data, in the beginning you may need to provide financial services without a trained algorithm, and train it using bad outcomes over time as they arise. This requires capital (to absorb the cost of these bad outcomes) and time, which raises the barriers to entry. Providers also mentioned that start-ups relying on new data sources are difficult to scale due to skills shortages, the difficulty of building relationships with data owners and the commercialisation of existing data and analytics solutions.

**Fear around data security and customer data protection.** Adoption of new tools or technologies in the use of data and analytics could be hampered by concerns around security. For instance, banks are cautious to access cloud-based data services, despite the benefit that can be derived from sourcing out analytics or certain data functions to specialist firms. Concerns about security are well grounded as consumer adoption can depend on the perception of personal data security. However, some organisations mentioned that security and privacy issues related to data are becoming easier to address through new practices. For example:

- **The tokenisation of data:** In an effort to increase the security of mobile payment systems, companies are starting to replace the confidential data of consumers with meaningless tokens.
- **Cyber insurance** is an insurance product covering Internet-based risks, which could include data related risks. As cyber security becomes increasingly important, cyber insurance will become increasingly important. This added security will enable organisations to get more involved in big data.

**Existing regulation cannot keep up with high speed of fintech development.** Regulation on the use of data and personal information has been adopted to various degrees in many countries. This includes consumer protection, data ownership, privacy, sharing, the cross border transfer of data and security. However, these regulations often rely on outdated principles focused on informed consent and choice for consumers. In an environment where the increasing availability and use of data makes it virtually impossible for consumers to read through every user agreement, users are effectively not informed and the current regulations therefore “[do] little to protect individuals today, or to support valuable new uses of data” (Cate, et al., 2014). Furthermore, early research shows that customers are often willing to sacrifice confidentiality and privacy for access to financial services (Chan & Faz, 2015). Some proponents are therefore suggesting a shift in regulatory focus to the responsible use of data, rather than the current focus on data collection and the existing notice and consent requirements (Costa, et al., 2015). While there is therefore scope to engage regulators on this topic, a lack of existing regulation on certain components of the use of data, like the cross border transfer of data, could make certain markets more favourable for new models of provision using data and analytics. Similarly, in markets where regulation is excessively restrictive or where there is a lack of clarity on the regulator’s stance, innovation can be stifled as providers tend to act conservatively when faced with regulatory uncertainty.

## Going forward

The **trends in data and analytics** identified above are likely to increase and will result in the increased adoption of alternative data and new analytical methodologies. However, the remainder of the note highlighted that there are opportunities to increase the speed of adoption, as well as enhance its potential to create new markets and improve the consumer experience for financial services. The following highlights some of the key issues that will need to be explored going forward.

Relevance for financial inclusion is not a foregone conclusion. Given the large proportion of the financially excluded or underserved that do not have traditional data footprints, the opportunities that the use of new data sources and analytics offer for expanding financial inclusion is an important consideration. Current trends in the adoption of mobile and smart phones are encouraging, but the question remains whether everyone will eventually benefit from new models for financial decision making and delivery. Adults that do not use technology or do not generate digital or alternative data footprints may still remain excluded. If so, the use of these models may play a role in widening the gap between the excluded and included by improving and increasing services only to those that have a data footprint. New data and analytical methodologies could also contribute to greater exclusion rather than inclusion if FSPs use new sources of data and innovative analytics to only pick the most profitable or least risky segments for provision.

<sup>7</sup> The Open Data Index is a global country survey with an indexed measure of Open Data available at <http://global.census.okfn.org/>

<sup>8</sup> Especially for algorithms that rely on feedback loops to improve outcomes, and the cost of replication is only low for the organisation that owns the algorithm. However, operationalisation and training new users is still difficult.

<sup>9</sup> However, start-ups can also form partnerships with existing providers which already have historical data on performance

**To drive adoption will require collaboration amongst the industry around key issues.** In order to ensure that the benefits of the use of new data and analytics are shared by as broad a group as possible, the industry has to adopt new sources and methodologies on a larger scale. This will require changing mind-sets, increasing investment, new partnerships and finding ways to scale new models quicker. Collaboration and discussion amongst a broad set of stakeholders, including governments, industry and the development community is therefore required. The following are some of the areas that would assist greater adoption:

- There is scope to increase **awareness, understanding and demonstration cases** for the use of alternative data sources and analytics for financial services provision. This will require information and evidence on the use and effectiveness of new models. This includes evidence on improved default rates, cost-benefit analysis and research geared to addressing FSPs' reservations and questions. In addition, there is a need to document more evidence based on successes and failures using alternative data and new analytical methodologies to reduce the learning curve for others interested in entering the space. Validating the potential of new models could be further fast-tracked by enabling and reducing the risk of experimentation through providing more information, facilitating more conducive policies and regulation and through direct investment.

Implementation and experimentation can be increased by addressing the strategic focus of organisations (shifting from a short-term focus to a longer-term strategy around the use of data) and experimenting with peer learning opportunities.

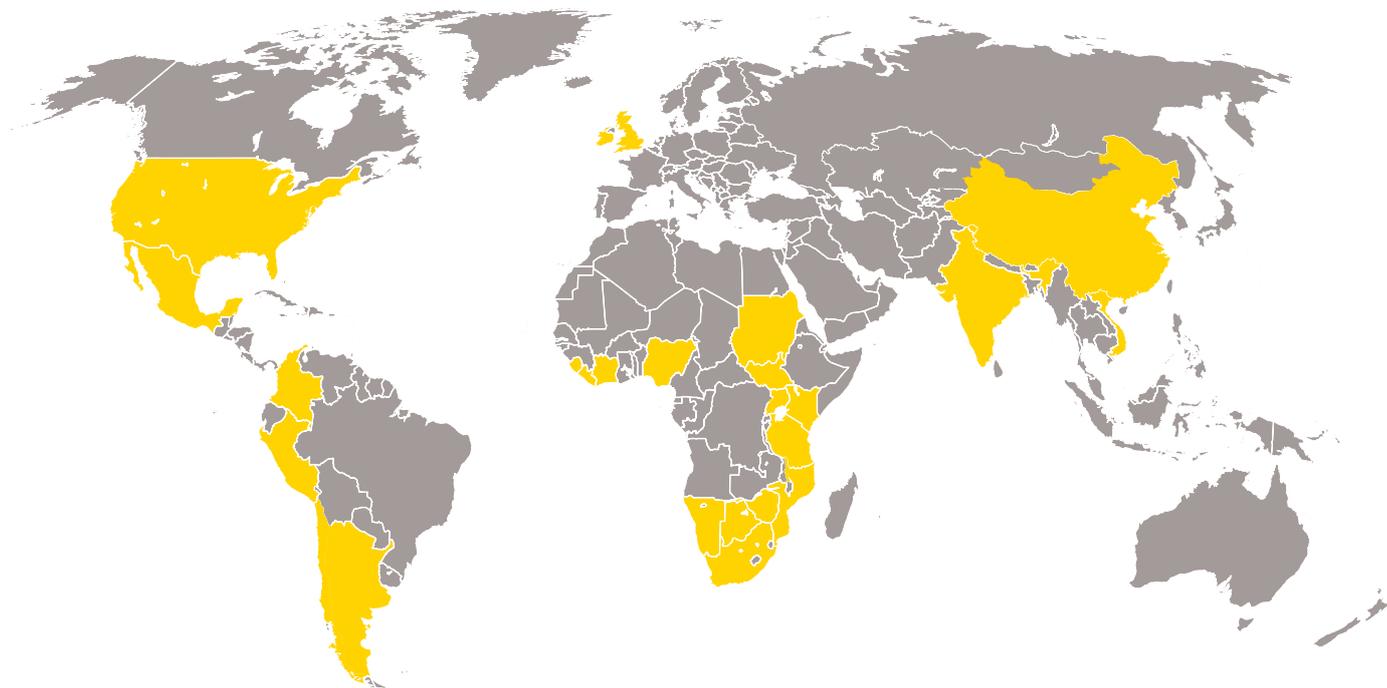
- Given the skills constraints identified, there is a need for **skills development and capacity building** to manage and make use of new data sources. This is required both within FSPs as well as in the external ecosystem that feeds in to the use of data and analytics by FSPs. This includes more capacity in local data experts that understand FSPs' needs and can work hand-in-hand with FSPs to optimise their use of data.
- The adoption of alternative data sources and advanced analytical methodologies, especially by larger organisations, will rely to some degree on internal structural changes and organisational change management which are needed to effectively internalise and embed new approaches within existing structures and systems. To ensure more effective adoption of new models, greater awareness of the **role of strategy and leadership** in this space is also needed, as well as assistance in creating and embedding a data strategy and ensuring that data leadership and data-driven decision making permeate throughout the organisation.

- Last, there is a need for more discussion and encouragement on **data sharing, data sharing principles and best practices** to facilitate greater access to data and reduce the friction associated with data integration. For example agreement on naming conventions would help reduce the friction of data integration. A better understanding of the incentives of different stakeholders involved in data sharing would be helpful and in some cases, government intervention to ensure greater access to industry data may be required.

**Emerging risks need to be managed to enhance value for, but also protect customers.** The benefits of wider use of new data sources and analytics will not come without risks. Protection and regulation will be increasingly important to ensure that consumers have adequate control over their personal data and to protect them against abuse and unethical conduct. There is therefore a general need for policymakers and industry to work together to rethink regulation relating to the use of data. Given the capabilities that new data sources and analytics unlock for FSPs, discriminatory practices may take hold – knowingly or unknowingly. In some countries, using particular variables to influence the price of and access to a product is already illegal. For example, in the UK insurance markets are not allowed to discriminate based on gender. In others this type of regulation may not exist. The challenge ahead will be to create an environment that allows for progress and innovation in the use of client data and analytics in a way that would enhance value for, but also protect customers.

# Appendix: Organisations

Geographies covered for the purpose of this study



Organisations		
22seven	Enigma	Musoni
ABSA	FirstAccess	Nirvoy Life Insurance
BancABC	IBM Analytics	Revolution Credit
Barclays	iHub	Saida
Beyond the Arc	inVenture	Segovia
Branch	Jumo	Social Intelligence
Brighterion	KCB	Social Lender
Capitec	KCB M-pesa	Superfluid Labs
Cignifi	Lenndo	Tiixa
DemystData	Leyebao Alifinance	Umati Capital
EcoCash Loans	M-KOPA	USC Data Science Lab
Effective Intelligence	Moven	Zest finance
EFL	M-Shwari	Zoona

# Bibliography

- Aglionby, J., 2016. Fintech takes off in Africa as lenders tap into mobile technology. Financial Times , 17 May .
- Byun, J., 2015. Zoono: A Case Study on Third Party Innovation in Digital Finance, s.l.: FSD Zambia .
- C3IT, 2016. Software as a Service: Why Should I care About the Cloud?, s.l.: s.n.
- Cate, F. H., Cullen, P. & Mayer-Schonberger, V., 2014. Data Protection Principles for the 21st Century , s.l.: Oxford Internet Institute.
- Cenfri, 2016. Making Access Possible (MAP) Zimbabwe: Financial Inclusion Country Report, Cape Town : The Centre for Financial Regulation.
- Chan, G. & Faz, X., 2015. The potential of digital data: How far can it advance financial inclusion? , s.l.: CGAP.
- Constine, J., 2014. Facebook Hits 100 Million Users in Africa, Half the Continent's Internet-Connected Population, s.l.: Tech Crunch.
- Cook, T. & McKay, C., 2015. How M-Shwari works: the story so far, s.l.: CGAP and FSD Kenya.
- Cook, T. & McKay, C., 2015. Top 10 things to know about M-Shwari. [Online]  
Available at: [www.cgap.org](http://www.cgap.org)
- Costa, A., Deb, A. & Kubzansky, M., 2015. Big Data, Small Credit: The digital revolution and its impact on emerging market consumers, s.l.: Omidyar Network.
- Crowley, A., 2015. Microcredit Brief: South Africa's Jumo launches microcredit borrower assessment technology based on mobile phone usage data, s.l.: MicroCapital.
- Davis, T., 2013. Open Data Barometer: 2013 Global Report, s.l.: Open Data Institute and World Wide Web Foundation.
- Entrepreneurial Finance Lab, 2016. EFL's Official Website, s.l.: s.n.
- Facebook, 2014. Connecting 100 Million People in Africa , s.l.: Facebook for Business .
- Far, X. & Chan, G., 2016. The Open Data Index. [Online]  
Available at: <http://global.census.okfn.org/>  
[Accessed July 2016].
- Financial Inclusion Insights , 2015. Value added financial services in Kenya: M-Shwari, s.l.: InterMedia Corporation .
- Gantz, J. & Reinsel, D., 2010. The Digital Universe Decade- Are You ready?, s.l.: International Data Corporation.
- Groenfeldt, T., 2015. Lenddo Creates Credit Scores Using Social Media. Digitalist Magazine , 5 February .
- GSMA, 2015. State of Industry: Mobile money. [Online]  
Available at: [http://www.gsmamobileeconomyafrica.com/GSMA\\_ME\\_SubSaharanAfrica\\_Web\\_Singles.pdf](http://www.gsmamobileeconomyafrica.com/GSMA_ME_SubSaharanAfrica_Web_Singles.pdf)  
[Accessed 12 July 2016].
- GSMA, 2015. The Mobile Economy: Sub-Saharan Africa, s.l.: GSMA.
- IBM, 2016. IBM. [Online]  
Available at: <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>  
[Accessed 12 July 2016].
- IDC, 2014. The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things, s.l.: International Data Corporation.

Internet World Stats, 2015. Africa Internet Users November 2015, s.l.: Internet World Stats.

Jamal, F., Luft, J., Shakhovskoy, M. & Wallace, S., 2016. The Rise of the Data Scientist: How big data and data science are changing smallholder finance, s.l.: The Initiative for Smallholder Finance.

Kelly, H., 2012. Lenddo Uses Social Networks to Give Loans and Improve Credit , s.l.: Venture Beat .

Leganza, G. & Angel, N., 2016. What You Need to Go From Data-Rich to Insights-Driven: Becoming Customer-Obsessed Requires People, Process, Technology, and Data Changes , s.l.: Forrester .

Macmillan, R., 2016. Digital Financial Services: Regulating for Financial Inclusion , s.l.: ICT.

McKay, C., 2015. Emerging Applications of Data Analytics: The case of M-Shwari. s.l., CGAP.

Musarirwa, T., 2016. Econet Not Suprised by the \$105.2 Million Revenue Decline. Techno Mag.

Norton Rose Fulbright, 2014. Global Data Privacy Directory, s.l.: Norton Rose Fulbright .

Out-law.com, 2016. Open APIs for UK Banking: It's Happening, People, s.l.: The Register .

Pelto, T., 2013. Is Social Media Impacting Your Credit? , s.l.: Quicken Loans .

Raj, V., 2015. Financial Inclusion Trends and Innovators-2015, s.l.: The Centre for Financial Inclusion & Accion .

Social Lender , 2016 . How Social Lender Works , s.l.: Social Lender .

TransUnion & Versta Research, 2015. The State of Alternative Data: Benchmark survey of 300+ lenders reveals insights into the adoption and competitive impact of alterntive data, s.l.: TransUnion .

World Bank, 2015. Private Credit Bureau Coverage in Sub-Saharan Africa, s.l.: Trading Economics.

# How to find us

Get involved. Contact us.

## Mari-Lise du Preez

T +27 (0)21 913 9510  
E [mari-lise@i2ifacility.org](mailto:mari-lise@i2ifacility.org)

## Celina Lee

T +27 (0)21 913 9510  
E [celina@i2ifacility.org](mailto:celina@i2ifacility.org)

## i2ifacility.org

-  @i2ifacility
-  /insight2impact
-  /insight2impact
-  /i2ifacility

Established by



Sponsored by



BILL & MELINDA  
GATES foundation