GEODATA FOR INCLUSIVE FINANCE: DATA PRIVACY

September 2019
By: Ben Wallingford and Kieya Derrick

Make the Difference
We Prove It

COLOPHON

Funded by:
NpM, Platform for Inclusive Finance

With Support From:
The Smart Campaign

Conducted by:
Microfinanza Rating (MFR)

Authors:
Ben Wallingford (MFR)
Kieya Derrick (MFR)

Co-Readers:
Mrinal Verma (NpM)
Alexandra Rizzi (SMART Campaign)

ACRONYMS

FSP	Financial Service Provider
GPS	Global Positioning System
GDP	Gross Domestic Product
Fintech	Financial Technology
IoT	Internet of Things

Cover photo by Mrinal Verma

TABLE OF CONTENTS

Photo by Marchmont Communications

EXECUTIVE SUMMARY

Financial institutions have historically stifled credit to the agricultural sector, citing high risk and lack of sufficient data to make an informed judgement on the credit worthiness of farmers. However, with the aid of technology, FSPs have begun to implement new innovative techniques to collect and analyze alternative forms of data to assess farmers with no traditional financial records. Among these innovations is the use of geodata which includes information on farmers' addresses, property locations, farm satellite imagery, as well as crop yield. This data can be used to estimate a farmer's income and their ability to repay their loans; it does, however, involve the disclosure of personal information and data, which raises a question as to what measures should be put place to ensure such sensitive information is well protected and responsibly used.

Comprehensive data privacy and protection standards may not be well established in many countries where personal geodata is used. FSPs are therefore tasked with the responsibility of safeguarding their clients' personal information by adopting best practices, especially where the data is shared with third parties (who may not necessarily be bound by adequate client protection standards). The main good practices identified in this study are robust internal policies, transparency, purpose/storage limitations, data security, accountability, and data confidentiality.

However, as is the case for many disruptive innovations, FSPs are posed with challenges in implementation of these best practices given the dynamic nature of the data collection and analysis process and the many players involved. The main challenges identified in this study are a lack of supervisory/regulatory guidelines, lack of client awareness, demographically identifiable data, discrimination by collective association, open data platforms, re-identification techniques, restriction on third party sharing, conflict of intentions, and high cost.

INTRODUCTION

In coordination with NpM, MFR conducted research on privacy risks in the use of geodata in digital financial services for agriculture. The research draws on MFR's own experience in client data privacy, insight gained from case studies conducted (including interviews and review of documents) on FSPs which have integrated the use of geodata in their agriculture loan products, and desk research.

This report contains key issues identified regarding collection, storage, use, and sharing of client geodata with a focus on best practices, challenges, as well as arising open questions.

MFR acknowledges the participation of the following Kenya-based institutions for the information provided and the time resources allocated for the purpose of this study:

- **Agri-wallet**: a digital platform that facilitates value-chain financing between farmers, suppliers and buyers through earmarked funds that are tokenized ensuring retention of funds on the agricultural value chain.
- **Tulaa**: a digital platform that links input suppliers, farmers, and commodity off-takers. It also provides financing to farmers for agri-input purchases and coordinates their delivery through existing retail networks.

MFR also acknowledges the support and advice provided by the Smart Campaign during the production of this report.

CONTEXT

Agriculture is the largest sector in Kenya. Contributing approximately 26%[1] to the country's GDP, the sector is fundamental to Kenya's economy. Approximately 38%[2] of the total population and more than 70%[2] of Kenya's rural people are engaged in agriculture.
Despite its importance to the economy, investment activity in the sector is underwhelming. Although many commercial banks and microfinance institutions in Kenya have rolled out a variety of products specifically tailored to agribusiness, the share of agriculture financing as a proportion of total outstanding credit remains below 5%[3].

Financial institutions have limited their exposure to the agricultural sector, which mainly comprises smallholder farmers[4] due to several factors, many of which are tied to data limitation. These include:

- Farmers in rural areas have limited access to formal financial services and as a result, have no credible and verifiable financial data to support lending decisions;
- Lack of collateral that could secure their loans;
- High costs of reaching remote rural populations;
- Risk of credit losses due to various factors such as loan misuse;
- Overall risk of the agricultural sector, i.e. production, climate, weather, and price risks.

The rapid expansion of Fintech platforms and the high penetration of mobile money in Kenya has enhanced rural outreach and proved to be key catalysts for FSPs as they attempt to bridge this financing gap to the agricultural sector. However, data on credit worthiness and overall risk profile of farmers remains a challenge. To address this, FSPs have adopted new innovations that enable the collection and analysis of alternative data for clients with no established credit record. Key among them is the use of geospatial information or geodata[5]; Institutions are testing how geographic data can be converted into useful information for credit scoring models to predictively assess a farmer's creditworthiness.

Geodata comes in various forms (e.g. GPS co-ordinates, altitude, climate characteristics, satellite imagery etc.) which provide decision points to drive the credit decision process. These parameters include:

1. Data from the Kenya National Bureau of Statistics (KNBS) indicates that agriculture and forestry contributed to 26.3% of the country's GDP in the first quarter of 2019 and 21.3% in the year 2018.
2. Food and Agricultural Organisation (FAO) 2016 country snapshot- Kenya
3. According to the Central Bank of Kenya's (CBK) quarterly statistical bulletin (June 2019) agriculture accounted for 2.3% of outstanding credit.
4. The World Bank's CGAP defines smallholder farmers as farmers that work a plot of land no larger than 1 hectare. The Kenyan Ministry of Agriculture estimates that smallholder farmers made up 70%-80% of farming actors, however no official research has been conducted to verify this estimate
5. Defined by ISO/TC 211 as any data with implicit or explicit reference to a location relative to the Earth

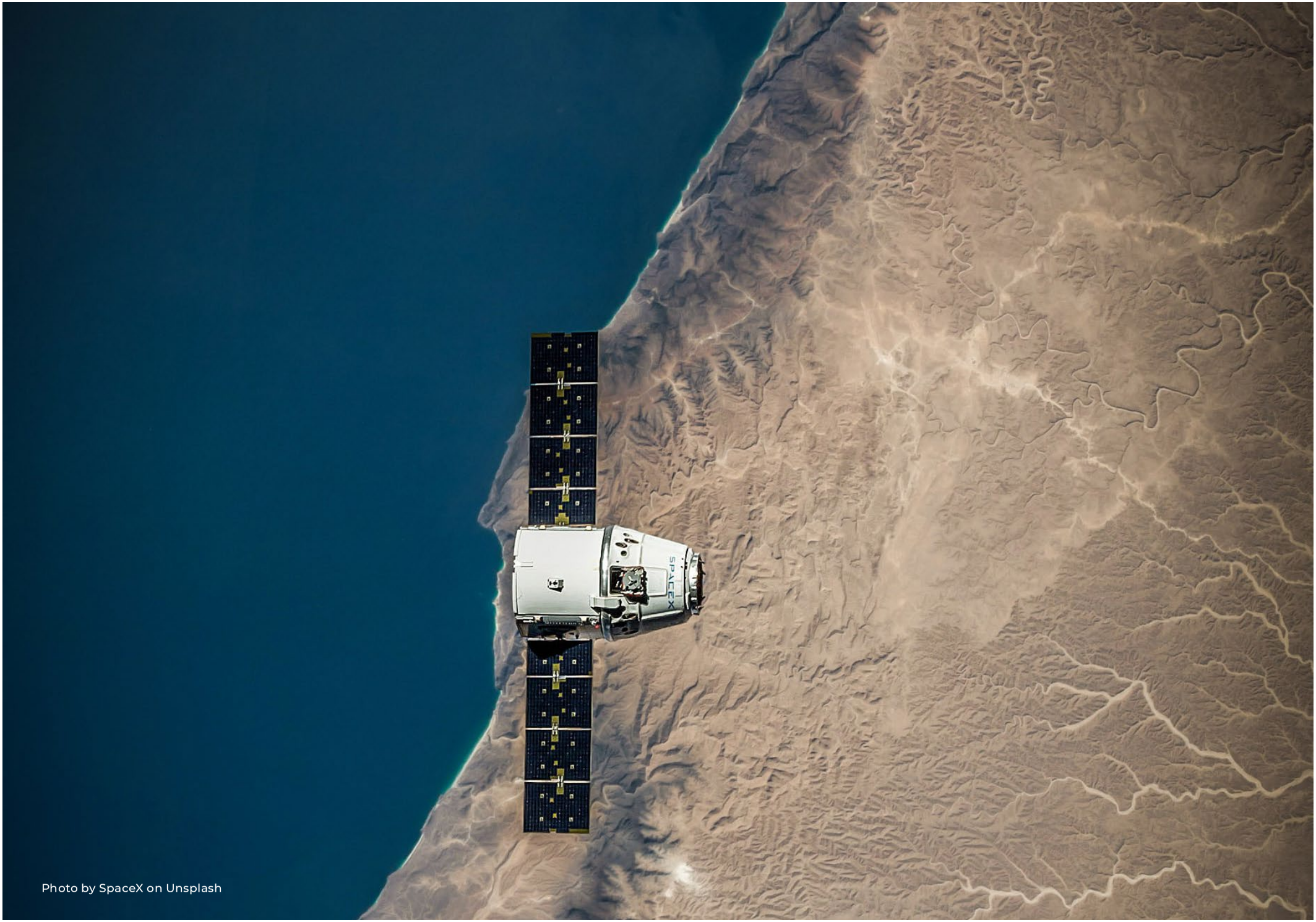

Photo by SpaceX on Unsplash

- Measuring land size and geotagging farmland to specific farmers - This enables the FSP to estimate the yield for each individual farmer based on land size and, in turn, estimate revenues. Geotagging prevents fraud, where multiple individuals take up loans on the same parcel of land;

- Identifying weather and climate patterns as well as planting cycles - FSPs are able to determine the suitability of a particular region for production of specific agricultural products. Repayment schedules for clients can also be set to match harvesting cycles;

- Analysis of historical farm crop yields and trends in production - Historical data enables FSPs to estimate future yields, and therefore revenues;

- The spread of crop disease in a particular area (probability/severity) - FSPs are able to factor in an additional risk factor in terms of crop disease when evaluating clients in particular areas. In some cases, they are able to offer their clients crop insurance;

- Mapping clients to the market characteristics of their region (prices, demand) which better informs revenue estimates.

This information can help forecast revenues, potential repayment deficits, and timing of income. Satellite imagery also enables potential post disbursement monitoring of the activities carried out on the farm.

Despite its advantages, the use of geodata also presents ethical dilemmas such as privacy and security concerns. While the majority of geodata is publicly available and is less sensitive on its own, it can be cross-referenced and coupled with additional information to form a personally identifiable data set. For example, most satellite imagery is openly available data; however, when a specific individual is tagged through GPS by their home location and the details of land they own, then the resultant data set becomes private and sensitive information. The industry must, therefore, define the required measures to ensure clients' interests are protected.

This report addresses this issue by considering existing regulation, identifying good practices, challenges, and open questions in regard to ensuring privacy in use of geodata. ■

REGULATORY BACKGROUND

Kenya has neither specific legislation on data protection nor an established data protection agency or authority. However, law makers have made significant strides towards ensuring the protection of personal data and defining rights with regards to the privacy of personal data. Most notable is the Data Protection Bill which was introduced in July 2018, and is currently in parliament. The principal objective of the Bill is to establish a legal framework for the protection of a person's privacy in instances where personal information is collected, stored, or used by another party[6]. Although the Bill has no specific provisions on geodata, it provides guidelines on overall personal data privacy, with the following key policies:

- Establishment of an independent authority to regulate data collection and register all institutions which collect and process personal data;
- Ensuring that information is collected directly from the data subject[7] and with their consent;
- Ensuring that collected information may only be released to a third party or put to a different use (to what was originally communicated) with the consent of the data subject;
- Ensuring that the data subject is informed (at the time of collection) of the purpose of the data collection exercise and the intended recipients of the information;
- Information should not be kept for a longer duration than is necessary for achieving the purpose for which it was collected;
- Reasonable steps are taken to ensure that the information collected is accurate, up-to date and complete;
- Appropriate technical organizational measures should be taken to safeguard the data subject against the risk of loss, damage, destruction of or unauthorized access to their personal information;
- Data subjects have a right of access to their personal information and a right to demand correction if the information is inaccurate.

Although the Bill covers key issues that are inherently applicable, some aspects of geodata present significant challenges to regulation. As is the case with many technical disruptions, the law is bound to struggle to keep up with its fast pace of change. This is coupled with the fact that the exchange of data extends across international boundaries making it even more difficult to manage.

6. The guidelines are applicable to any natural or legal person, public authority, agency or other body involved in collection and processing of personal data.
7. Defined in the Bill as an identified or identifiable natural person who is the subject of personal data.

GEODATA PRIVACY: GOOD PRACTICES

Best practices around client data privacy involve careful consideration of the type of data required, the process of collecting the data, the use, storage, and sharing of this data. The Smart Campaign's 7 client protection principles[8] for responsible financial inclusion are a key reference point as a framework for identifying adapted good practices, in particular the principle on 'privacy of client data'. This principle provides guidelines for FSPs on client data privacy rights and how to appropriately handle client data.

The following are some of the best practices observed from MFR's own experience in data privacy, adaptation from the Smart Campaign's client protection guidelines, case studies of FSPs (including interviews and review of documents), and desk research.

ROBUST INTERNAL POLICIES

Before commencing any geodata collection exercise, institutions should ensure that relevant policies on handling of the data have been formalized. These include a clear definition of the data to be collected, the channels to be used, ownership and access to the data, and the sharing of data with outside parties. More specific internal best practices include:

- Preparing and maintaining an up-to-date privacy risk assessment to document the personal information used, identify privacy risks to clients during collection/processing/storage/sharing of data, and put in place mitigations as needed;
- Adopting and implementing comprehensive data protection policies and procedures to protect against misuse of geodata;
- Appointment of a data protection officer to oversee general data privacy and security;
- Having contracts with third parties with access to data in order to ensure that obligations for safeguarding client data are aligned;
- Keeping records of data processing, storage, and sharing activities;
- Training of staff on data protection practices.

TRANSPARENCY

The institution should be transparent with the individuals whose geodata is being collected and ensure consent is obtained when feasible. Additionally,

8. Appropriate product design and delivery, Prevention of over-indebtedness, Transparency, Responsible pricing, Fair and respectful treatment of clients, Privacy of client data, and Mechanisms for complaint resolution

some basic level of client awareness training should be conducted to inform unknowing individuals about the nature of the data collection exercise, the purpose for which the data is being collected, and the rights the individuals have over the sharing and use of their personal information. It should be made clear to farmers if the contract involves sharing with third parties, business partners, or affiliates. The terms and conditions of the agreements should be in clear language and/or translated where possible.

PURPOSE/STORAGE LIMITATIONS

The use of geodata in credit appraisal is still highly experimental and no standard models have been developed to determine the ideal set of data points that best estimate a farmer's repayment capacity. As such, FSPs tend to collect large volumes of client information with a 'lend-to-learn' approach. FSPs should therefore constantly evaluate their algorithms and develop a better understanding of which sets of geodata are significant in determining farmers' payment behavior. The objective of this process is to limit the amount and detail of client information collected (especially personal/identifiable client information and sensitive types of geodata) to what is absolutely necessary. 'End-of-life' policies should be in place and implemented for data that is deemed redundant. Where applicable, farmers should have the right to opt-out of providing a particular type of sensitive geodata, while still being eligible to proceed with loan application.

DATA SECURITY

FSPs should adopt best practices for data security by implementing access controls, systems monitoring, and proper information risk management. In the event of a data breach, FSPs should be obligated to notify relevant supervisory authorities and clients within a reasonable period. An impact assessment should be conducted to identify the nature of the breach, affected parties, and the likely consequences. Measures to be taken to address the breach and mitigate the adverse effects should also be communicated to all affected parties. In the case that clients use a mobile application to access products and services, clients should be notified by FSPs about available application updates that are important for their security and privacy.

ACCOUNTABILITY

Agreements with third parties accessing geodata should ensure the same privacy standards of the FSP are maintained, including commitments to using the data exclusively for the intended purposes, ensuring data security, and prohibiting further sharing. Penalties for breaching the agreement should also be included in the contract.

DATA CONFIDENTIALITY

FSPs should aim to limit the amount of individually identifiable information collected. However, some personal geodata is relevant to the credit approval

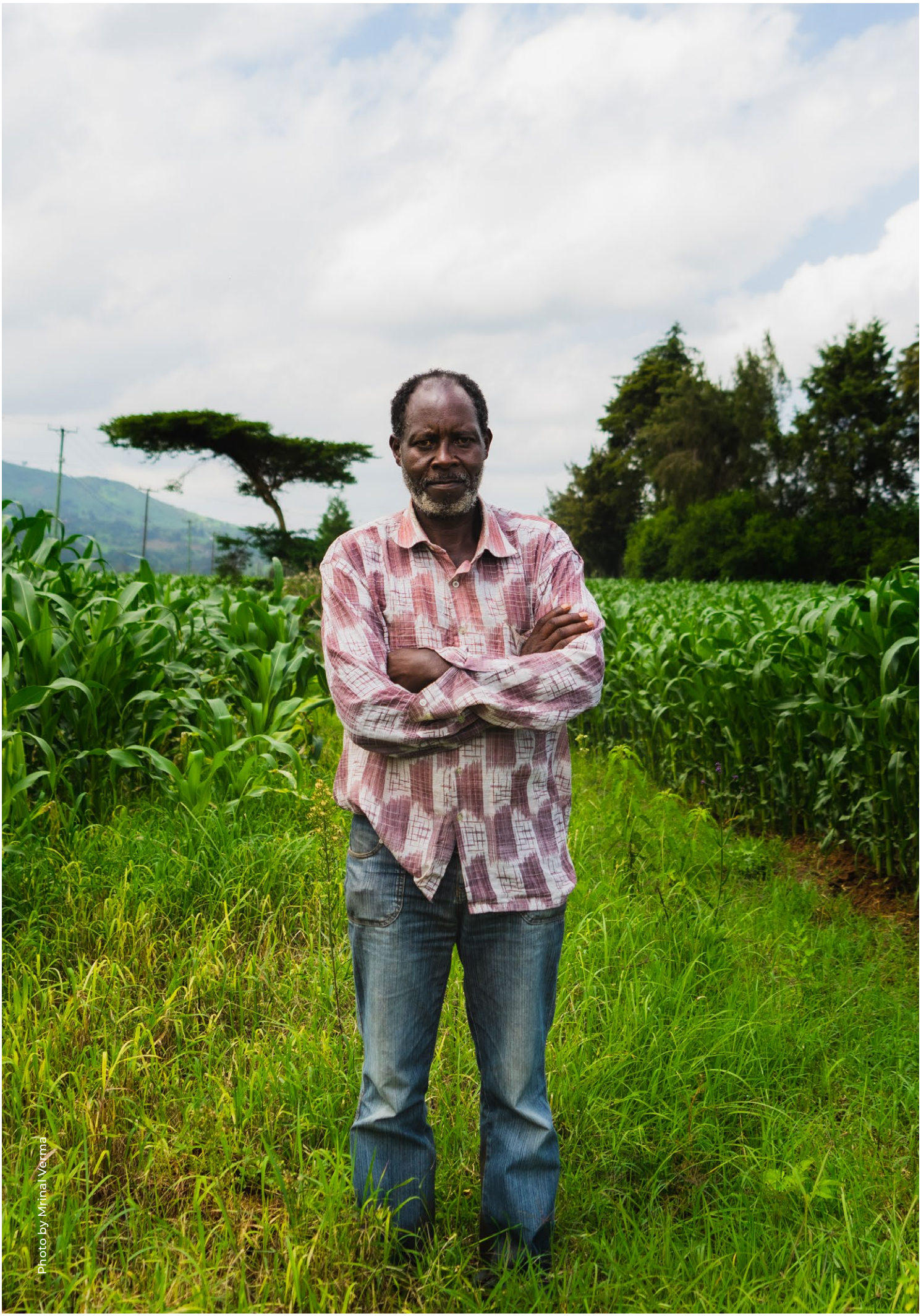

Photo by Mrinal Verma

process and cannot be left out; Such data should be obscured using anonymization[9] methods such as:

- Anonymizing data for staff or third parties who don't need the information, with internal access rights controls in place, especially for sensitive types of geodata;
- Geomasking techniques which alter data to minimize identification of individuals while still maintaining relevant characteristics. (E.g. Location swapping: where an individual's originally recorded co-ordinates are swapped with a masked location selected from a set of possible locations with similar geographic characteristics within the same neighborhood)[10];
- Using satellite imagery at the lowest practical resolution and removal of identifying visual characteristics while ensuring the data remains useful.

9. Defined in the Kenya Data Privacy Act as the irreversible removal of personal identifiers from personal data so that the data subject is no longer identifiable.
10. Su Zhang, Scott M. Freundschuh, Kate Lenzer & Paul A. Zandbergen (2017). The Location Swapping Method for Geomasking, Cartography and Geographic Information Science.

GEODATA PRIVACY: CHALLENGES

LACK OF SUPERVISORY/REGULATORY GUIDELINES

The absence of specific guidelines in regard to geodata means that there are no standard definitions or procedures to be uniformly adopted. In Kenya, sensitive personal data refers to any data that reveals a person's race, health status, ethnic social origin, political opinion, belief, personal preferences, location, genetic data, biometrics, sex life/sexual orientation, or personal financial expenditures[11]. In the case of agriculture loan clients, additional information collected such as land location, satellite imagery, crop yield data, farm input preferences, etc. does not fall into the usual categories of personal information that companies are legally obligated to protect. It is therefore at the FSP's interpretation as to whether or not such information should be considered sensitive.

LACK OF CLIENT AWARENESS

The majority of clients are marginalized and are often not aware of the sensitive nature of the geodata they disclose or their rights in regard to handling of the data. Clients are therefore not in a position to make any legitimate queries, identify errors, or report any violation of data privacy (intentional or otherwise). Client feedback is therefore lacking in this regard, which is instrumental to helping institutions improve their internal processes.

DEMOGRAPHICALLY IDENTIFIABLE DATA

Demographically Identifiable Data is defined as data points (individual or aggregated) that allow inferences to be drawn that enable classification or identification according to ethnicity, economic class, religion, gender, age, health condition, location, occupation, or other demographically defining factors[12]. Even when personal geodata is de-identified, clients still face the risk of stigma and discrimination resulting from being associated with particular locations; data may not personally identify an individual but it enables them to be tracked or classified to the factors above[13].

DISCRIMINATION BY COLLECTIVE ASSOCIATION

Geodata identifies trends in locations. However, general trends may not be accurate representations of all individual cases that may be affected by decisions made based on them. Clients therefore face the possibility of being assigned certain attributes erroneously, as a result of being collectively associated with particular geographical areas. For example, a case where one farmer's inability to repay their loans has been directly factored in when assessing subsequent loan applications for other farmers in the neighboring area – a default by one individual increases the risk associated with other individuals of the same group.

11. According to the Kenya Data Privacy Act.
12. Raymond, Nathaniel. (2017). Beyond "Do No Harm" and Individual Consent: Reckoning with the Emerging Ethical Challenges of Civil Society's Use of Data. 10.1007/978-3-319-46608-8_4.
13. Ethical Considerations When Using Geospatial Technologies for Evidence Generation. UNICEF Office of Research - Innocenti Discussion Paper DP-2018-02 | June 2018.

Photo: Alamy Stock

OPEN DATA PLATFORMS

Technological advancements and regulation (both passively and actively) has made data collection easier and more intrusive. Some types of geodata are already publicly available and do not require any consent (e.g. satellite imagery). With over 1,000 satellites in orbit and more than 15 platforms offering open data resources for satellite imagery and other geographic information, it may not be realistic to expect consent for such information. In addition, other forms of geodata are obtained from alternative sources such as social media or telecommunications and it would be hard to argue that consent for the particular use of this type of geodata (even aggregated in an anonymized way) is truly informed11.

RE-IDENTIFICATION TECHNIQUES

The large volume of geodata available means that even if obscured, it still bears the risk of being coupled together to form a set with identifiable characteristics. With additional datasets coming from smart phones, personal computers, and IoT devices, information can be collected and pieced together for any individual. This fundamentally changes the ability of anonymization methods to effectively protect peoples' privacy.

RESTRICTION ON THIRD PARTY SHARING

The lending process often requires partnership with third parties (e.g. mobile money providers, outsourced data processing services, debt collectors). In such cases, outside parties become privy to sensitive client information. While it is possible to mitigate incidents of privacy violation by having contracts in place, it becomes increasingly difficult to ensure that data is safeguarded when several players are involved; especially parties who do not necessarily share similar sentiments regarding client data privacy and may not be bound by standards of best practice. In addition, FSPs are not able to fully implement end-of-life protocols for information that is no longer required given that the data is collectively accessed.

CONFLICT OF INTENTIONS

While it would be best practice for FSPs to limit the amount (and level of detail) of information collected on clients, in most cases gaining more insight on client characteristics improves accuracy of the loan appraisal methodology. Thus, lies the dilemma for FSPs who wish to respect personal data privacy while at the same time require more personal data to form a better opinion on credit worthiness.

HIGH COST

Adhering to best data privacy practices requires resources that many FSPs do not have. The process involves procuring the required hardware and software, having specialized data protection personnel, training of staff and going the extra mile to educate clients. Many institutions may not consider this a priority given the cost involved. ■

14. E.g. In 2014 the US Department of Commerce relaxed its restriction on minimum resolution for commercially available satellite imagery. Leading Satellite imagery provider DigitalGlobe was given approval to sell its highest-resolution imagery (with detail of down to 25 cm) to all customers. This information was previously only accessible to the government.

GEODATA PRIVACY: OPEN QUESTIONS

Do the benefits of collecting geodata outweigh the privacy risks, keeping in mind the farmer's perspective?
FSPs mitigate risk of loss by factoring in as much information as possible when assessing an individual's credit worthiness. Given that not all this information may be relevant, it is important to assess whether the potential benefits of gathering this information outweigh potential risks of personal data privacy breach or other data privacy risks for their clients.

Is the concept of informed consent even more important for geodata, as compared to other types of personal client information? Are all types of geodata acceptable for use in inclusive finance if farmer consent is transparently gained?
Given the arguably intrusive nature of some types of geodata, consent may be even more important to receive in a fully informed way than other types of client information.

Are there some types of geodata that are 'too' intrusive? Which types of geodata are more intrusive than others?
Given the vast amount of data available, should there be a limit on how much detail should be captured on a client?

How best to gain consent of geodata use in a way that the farmer truly understands what is being consented to, including in a digital context?
As highlighted in the previous chapter, client awareness poses a major challenge. The majority of farmers are unaware of the nature, and specifically the sensitivity, of the information they disclose. They may not be aware of what they are giving consent to, and in most cases, they have no choice but to give consent due to lack of alternatives. FSPs should carefully consider investing in client education and other channels to enhance transparency.

Should farmers be able to opt out of sharing certain types of geodata and still be eligible for loan approval?
Different individual clients will tend to hold different degrees of reservation regarding what type of personal data they are comfortable sharing. FSPs should allow farmers some degree of freedom to withhold certain types of geodata without being disqualified from the loan application process, where possible.

What challenges do we see in terms of 'ownership' of geodata?
Taking into account the various open data sources available and the network of service providers (FSPs, payment platforms, telecommunication service providers etc.) with shared access to any particular set of collected data, it presents several gaps in regard to ownership of, and accountability for, the data.

Should GPS information be gathered only at the application stage, or on an ongoing basis?
Recording clients' residence and farm GPS co-ordinates is among the first steps during registration, it remains an issue whether the initial recording is sufficient or whether data should be continually updated. Although arguably intrusive, constantly updated GPS information is useful for loan use and post disbursement monitoring.

NEXT STEPS

The use of geodata has the potential to expand outreach of financial services to the agriculture sector at a massive scale and at significantly lower costs. This great potential is however coupled with an equally great risk of data privacy violation for uninformed farmers. In absence of comprehensive regulation, FSPs are tasked with exploring the advantages as well as acknowledging and actively mitigating the risks of collecting and analyzing geodata.

FSPs and investors can take steps now to facilitate identifying and implementing best practices standards in geodata:

- **Action by FSPs**: FSPs using geodata should identify client protection risks related to geodata use, and implement mitigations to achieve client protection. Being a relatively recent development, there are unfortunately limited resources to guide FSPs in the handling of geodata. However, there are existing tools available to inform FSPs on client protection and general client data privacy , including this report which is intended to provide further insight in this regard.

- **Support from Investors**: Investors should encourage FSPs and support them in adopting best practices. Investors should partner together toward the development of universally acceptable standards for handling geodata and push for widespread adherence of their investees to ensure client protection.

YOUR TAKE?

Opinions on the good practices, challenges, and open questions elaborated in this report are highly encouraged. Participation from a wide range of stakeholders will be key to further enhancing best practices and identifying additional ways to ensure responsible use of geodata. As such, interested parties are requested to contribute to this discussion by providing their input.

15. These include self-assessment tools and client protection guidelines developed by the Smart Campaign, and the Checklist for Ethical Use of Geospatial Technologies which is part of a research publication by UNICEF.


NpM
platform for
inclusive finance


MFR
Make the Difference
We Prove It