



Digital Credit Scoring in Agriculture:

Best Practices of Assessing Credit Risks in Value Chains



This report was sponsored by:



Strengthening Agricultural Finance in Rural Areas (SAFIRA) is a multi-year program under AIPRural aimed at developing and formalizing value chain financing with selected partners in eastern Indonesia to benefit smallholder farmers. Value chain financing is one way to scale-up cost effective, appropriate and sustainable lending to smallholder farmers without increasing transaction costs. SAFIRA partners with financial institutions, financial consultants, agri-businesses and farmer's groups to increase the incomes of up to 6,000 farmers. SAFIRA works in East Java, West Nusa Tenggara, East Nusa Tenggara and Papua. SAFIRA is part of PRISMA and is supported by the Australian Government and Bappenas. Implemented by Palladium with technical assistance of Swisscontact.



Grow Asia is a multi-stakeholder partnership platform that catalyzes action on inclusive and sustainable agricultural development in South East Asia. Grow Asia is part of the World Economic Forum's New Vision for Agriculture

Both SAFIRA and Grow Asia acknowledge the support of the Australian Government Department of Foreign Affairs and Trade:



Australian Government

Australia's aid program seeks to promote prosperity, reduce poverty and enhance stability. The program has a key focus on the Indo-Pacific region.

Contents

Introduction	8
Market Trends	11
Assessing Smallholder Credit Risk What Data is Best for Scoring Models?	12 13
Transaction Records Technological Approaches	14 16
Agronomic Survey Data Collected by Agribusinesses	17 18
Farmer Demographic Data	24
Satellite Imagery Data Weather Forecasts and Records	26 28
Credit History	30
Alternative Data Sources Mobile Phone and Wallet Usage Social Media and Networks Psychometrics	31 32 35 36
Putting it All Together - Which Data Sets are Complementary?	38
The Value of Digital Scoring to Agribusinesses	40
How to Engage with Financial Institutions Where to Start? Other Possibilities for Developing a Scorecard	42 42 43
The Importance of Loan Product Design Input Linked Loans Buyer Repayment Deduction	46 47 47
Risks and Challenges	48
Conclusions and Recommendations	49
Acknowledgements	51



Executive Summary

Farmer-level data collected from digital channels can be used to evaluate and rank potential borrowers in order of their likelihood to repay a loan. Such credit scoring models could help increase the scale and scope of lending to small holder farmers and augment existing value chain finance programs and informal lending.

Several concurrent industry trends are increasing the volume of digital data relevant to small holder farmers—from records of small holder farmer purchase and sales, the movement of goods through value chains (or ‘traceability’), and remote sensing of weather conditions, to farmers’ personal use of digital services on mobile phones and devices. Lenders and on-lenders (those in the value chain who borrow from the bank and in turn lend to small holder farmers) with access to such data may be able to develop credit scoring models to reasonably assess farmer credit risk at scale. This would allow lending to reach many more farmers than at present, while attempting to keep the risks and transaction costs low.

The value of each data set to a credit scoring model is a function of its availability from all farmers, relevance to farmer creditworthiness, cost to obtain, and reliability in predicting farmer credit risk. Ideally, a balanced scoring model would contain elements of credit history, transaction records, agronomic survey data and lifestyle-related demographics (marital status, household size, years in address, etc.), and could be augmented (or supplemented) by alternative data where feasible.

First digital scoring efforts may be wise to focus on the 40% of small holder farmers in structured cash crop value chains. The data on these farmers is likely to be more complete, and the farmers themselves more creditworthy. A phased approach to credit, starting with small, relatively simple and short-term products, can introduce farmers to the formal credit system and help them build credit history that unlocks larger, longer-term credit.

How to Read this Guide?

This guide reviews many types of digital data lenders can use to assess the creditworthiness of small holder farmers in cash crop value chains. It is written to be accessible to any reader, with its two main target audiences being:

- 1 Agribusinesses working with networks of small-holder farmers in structured value chains.
- 2 Financial institutions interested in lending to small-holder farmers at scale.

Following a brief introduction on current market trends and farmer credit assessment practices, the guide presents an overview of each of the following types of farmer and farm-relevant digital data sets being collected by various stakeholders for their own business reasons:

Agribusiness readers can learn how the transaction and survey data they gather to manage their farmer networks might also enable lenders to affordably assess and finance their customers. At the same time, they can understand how such data is similar, different, and potentially complimentary to other types of farmer data being collected by other businesses (such as weather services, credit bureaus, cooperatives, and alternative data aggregators).

Financial institution readers may be surprised to learn that much of the data they require for credit assessment—such as farm characteristics, historic transaction records, and forward-looking cash flow projections—are already being collected and/or prepared in digital format by agribusinesses. Other data sources offer lenders the potential to independently verify the suitability of land and weather conditions, identity (through social media), and borrower ‘character’ (credit bureau data or innovative psychometric tests).

The remaining sections of the guide look at how agribusinesses, financial institutions, ‘fintechs’, and other interested stakeholders can partner to explore what is needed—from the data itself, loan product design, potential risks and data privacy concerns—to begin using some of this data in credit scoring tools aimed at financing credit-worthy farmers.



Introduction

“Digital credit scoring” in the context of this guide refers to the use of farmer-level data shared over digital channels to evaluate and rank potential borrowers in order of their likelihood to repay a loan.

Digital credit scoring could potentially increase the scale and scope of lending to small holder farmers by complementing or augmenting existing value chain finance programs and informal lending from the farmer’s local community. Value chain finance programs lend based on the dynamics and relationships within specific structured value chains, rather than on data about those relationships. An excellent banker’s guide to value chain finance was recently published by AgriFin¹. Informal credit, the only choice for many small-holder farmers, is widely perceived to be expensive and a barrier to increasing small-holder farmers productivity.

This guide focuses the discussion of ‘first-mover’ digital scoring models in structured cash crop value chains (see box on structured value chains). The farmers in these chains make up roughly 40% of all small holder households worldwide, or 1 billion people in households². They are perhaps the best starting point for data-driven credit assessment because:

- Agri-buyers, particularly, are capturing more and more information about their network farms and farming behaviors in digital format. Dedicated software platforms help many agri-buyers digitize and manage small-holder farmer data, but such data can also be collected and digitized by farmer cooperatives, public-private partnerships, and capacity building programs
- Some of these agri-buyers have sophisticated traceability systems that use technology (i.e. bar codes, Radio Frequency Identification (RFID) and Near Field Communication (NFC)³.) to track produce from farm to end processor. The reasons for tracing agriculture produce include supply chain management, food safety, product certification, and ethical concerns
- The farmers in structured chains are more likely to have access to improved seeds, inputs, and secured markets.

Structure in value chains

Small-holder value chains derive structure from:

- The same buyers operating in the region year on year
- The farmer consistently growing the same crop
- Formally documented transactions with the farmer.

The diagram below presents value-chain structure on a spectrum from contract farming, the most structured arrangement, to basically unstructured ‘fragmented’ chains.



Beyond the value chain itself, wider trends in technology usage—from personal use of mobile phones and wallets, the internet, and social media to weather satellites and remote sensors—create ‘digital footprints’ of farmers that until recently were not even ‘on the map’ of most lenders (unless put there, at considerable cost, by personal visits to farms). The guide also looks at some other types of innovative data that can be collected from farmers remotely using digital devices.

Knowledge and Dialogue

Before starting this journey into the expanding universe of small-holder farm data, it is worth noting that this guide do not suggest that data, math, or technology alone will reveal a ‘universal’ model for lending to small-holder farmers. Digital data sets are unlikely, for example, to shed much light on the historical, socio-economic and regulatory issues that affect small-holder farmers and it differs by country. Instead, the guide aims to act as an informative resource for agribusiness value-chain actors, digitally-enabled service providers, and financial institutions (both traditional and ‘fintech’). Though no ‘silver bullet’⁴, digital scoring can certainly give lenders a promising, scalable starting point for reaching out to this traditionally challenging market segment.

¹ Please see AGRICULTURAL VALUE CHAIN FINANCE A GUIDE FOR BANKERS, 2016 World Bank Group <http://www.agrifin.org/sites/agrifin/files/microsite/doc/Bankers%20Guide%20to%20AVCF.pdf>

² CGAP Note “Segmentation of Smallholder Households: Meeting the Range of Financial Needs in Agricultural Families” (2013), p. 10.; <http://www.cgap.org/sites/default/files/Focus-Note-Segmentation-of-Smallholder-Households-April-2013.pdf>

³ Traceability across the Value Chain: Advanced tracking systems, Case study 40. European Commission Business Innovation Observatory Series. <https://www.google.co.uk/search?q=traceability+technology&aq=chrome..69i57.6252j0j8&sourceid=chrome&ie=UTF-8>

⁴The term ‘silver bullet’ to refer to an action which cuts through complexity and provides an immediate solution to a problem (<https://www.phrases.org.uk/meanings/silver-bullet.html>)

Market Trends

The prevalence of digital data in cash crop value chains should continue to grow in the coming years based on at least five long-term, consistent trends in the agribusiness industry:

- 1 Transaction Records**
 Agribusinesses (including input companies, end buyers and traders) are increasingly keeping digital records of their transactions with small-holders to improve the efficiency of their business operations. Efforts are also underway in many markets⁵ to encourage farmers to begin using mobile wallets (instead of cash) for purchases and sales.
- 2 Industry and Consumer Concern**
 Farmer's livelihoods has made tracking data on farms and the path of produce increasingly essential to sustainable business.
- 3 Digital Footprint**
 Farmers in some markets are beginning to use electronic services from a growing range of digitally-enabled service providers (i.e. payments made on digital platforms and use of mobile apps providing farm-relevant information). This trend is likely to continue to increase as the demographics of the farmers shift to a younger generation who are more digitally proficient.⁶
- 4 Remote Sensing**
 Improved satellite resolution and falling costs could facilitate the collection of highly relevant data (farm size, crop health) even without visiting a farm in person.
- 5 Improved Communications**
 It is increasingly possible to communicate with farmers by mobile phone to ask them for information.

More and better digital data can potentially help lenders to screen and assess small-holder farmers at scale (and thus at a reasonable cost). The next section discusses how lenders can use this data to assess credit risk.

⁵ Supported by the MasterCard Foundation and others, see, for example "Opportunities for Digital Financial Services in the Cocoa Value Chain, CÔTE D'IVOIRE" <http://www.ifc.org/wps/wcm/connect/2d3ae2fc-ae9a-45e1-bb9a-f039927a2f89/IFC+Cote+d%27Ivoire+Digitizing+Cocoa+Value+Chain+report+ENGLISH.pdf?MOD=AJPERES>

⁶ GSMA, Analysis: Mobile internet usage challenges in Asia – awareness, literacy and local content, <https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2015/07/150709-asia-local-content-final.pdf>

Assessing Small-holder Credit Risk

Financial institutions have at least three major concerns when they lend money:

- 1 **Verifying the borrower's identity (“Know Your Customer”)**
- 2 **Understanding how the borrower will earn money to pay back the loan**
- 3 **Mitigating loss if the borrower is unable to earn money as planned**

Small-holder farmers have traditionally been considered too risky for all but those financial institutions with a special mission to serve them. Farmers often have limited savings and assets, and their income and livelihoods depend heavily on the crops they plant and/or animals they raise. Success each year depends not only on timely receipt and planting of quality inputs, soil and vegetation conditions, and use of the appropriate fertilizers and pesticides, but also on things outside of their own control, such as rainfall and general weather conditions. At the same time, most lenders lack expertise in agronomy, leaving them with a complex risk-assessment equation in exchange for potentially small returns from each small-holder farmer.

Each digitized data set discussed in the sections below can contribute to a clearer picture of who the farmer is and how the farm will generate income. Data on past farm production, purchases, and sales can help answer a lender's questions about the farmer's **ability** to repay a loan. Mobile telephone and money usage patterns may shed light on the farmer's **willingness** to repay a loan⁷ (particularly in absence of formal credit history, which is the most direct measure of willingness to repay). Any other data indicating stability and track record is likely to be related to creditworthiness—for example, a farmer that has been farming his plot for 20 years has presumably managed to honor obligations to informal lenders in the past.

⁷ Mobile phone use patterns are quite predictive for repayment in consumer lending and we expect similar predictive power in business lending.

What Data is Best for Scoring Models?

While the reasons for and methods of collecting the various types of data differ, the potential contributions of each data type to a single credit scoring model are bound by a few basic truths about credit scoring models:

- 1 The more closely data is related to past income-earning activity and payment of obligations, the more likely it will be predictive of future loan repayment.
- 2 A model's risk ranking continues to improve as long as each new piece of data adds new information⁸.
- 3 The more objective and verifiable the data, the clearer its signal will be to a model (i.e. think of the difference between asking someone how much they earn or seeing the bank account their salary is deposited to).

This guide examines the following types of farm-relevant digital data sets collected by stakeholders:



To help readers consider the cost-benefit trade-offs of using various data sets for credit scoring, Chart 1 on page 39, which follows the presentation of all the data sets, summarizes the expected contributions of each type of data along the dimensions of Relevance, Availability, Cost to Lender, Reliability and Predictive Power.⁹

⁸ In mathematical terms, is not highly correlated with the other information in the model.

⁹ This is a general assessment based on the author's experience in general, but subject to variation in any particular situation

Transaction Records



Digital transaction records evidence business activity over time. In some cash crop supply chains, three actors collect and maintain farmer transaction data:

- **Traders** keep a record of what they buy from a farmer
- **Agro-input companies** keep a record of the products they sell farmers
- **End buyers** ask traders/aggregators for records of their purchases from farmers.

Each transaction record typically includes:

- Transaction value
- Transaction date
- Product quantity
- Product type
- Farmer's identity

Advances in technology have lowered the cost of capturing farm transaction data (although still relatively few supply chains have such traceability in place today). Using tablets, mobile phones, and computers, this data can be recorded at remote buying stations or even at the farm gate. Aided by more affordable devices, improving mobile coverage, and improved digital literacy, the volume of farmer transactional data should continue to grow in the foreseeable future.

In addition to lower data collection costs, several businesses need to drive the digitization of farmer transaction records including:

- **Securing Supply**
A range of factors from climate change to urbanization are tightening the supply of some small-holder crops including cocoa, coffee and palm oil. Knowing where products come from helps agribusinesses to better place buyer stations and to tailor and target training programs. Better understanding and supporting farmers lifts supply and establishes farmer loyalty, which is crucial to maintaining supply.

- **Food Safety**
This is particularly important for crops prone to pesticides residues. Digital transaction logs help track non-compliance shipments back to source. In many cases the transaction log is a regulatory requirement.
- **Ethical Concerns**
Consumers increasingly want assurance from Fast Moving Consumer Good companies (FMCGs) that the foods they consume are not driving deforestation or child labour and that farmers and their families are able to live in decent conditions. Transaction records let FMCGs know exactly where goods originate and facilitate product certification schemes (such as Rainforest Alliance and Fairtrade International).

Whichever the root driver for a given agribusiness, the result is that transaction logs are growing. While these logs are not usually collected for the purpose of digital credit scoring, they make an excellent source of data for lenders interested in building credit scores. The next section looks at the types of technology platforms currently storing transaction data.

Technological Approaches

Transaction records can be kept on the following technology platforms:

- 1 An internal database**
 This is the simplest form of logging, and may be done merely to provide the farmer with a receipt and meet basic tax and regulatory obligations. In the very simplest examples, buyers may keep digital records of transactions in an Excel-based ledger.
- 2 Traceability Software**
 A number of traceability software solutions have been developed specifically to record farmer data (vendors FarmForce, GeoTraceability, and Koltiva are profiled in call-out boxes). These software solutions record not only transaction data, but also a variable range of farm and farmer specific data, including basic demographics, GPS location, farm practices and training attendance.
- 3 Mobile Wallet**
 In markets with adequate mobile network coverage and mobile money penetration, some buyers may use mobile wallets to pay farmers for produce. Mobile payments are potentially more secure and less costly to deliver than cash payments. For buyers using mobile wallets, the buyer and wallet operator would keep these transaction records.
- 4 Blockchain**
 Isolated examples are emerging with buyers recording transaction data in a block chain ledger (such as Provenance, which traces yellowfin and skipjack tuna fish in Indonesia from catch to consumer¹⁰).

These four platforms house datasets useful to understanding a farmer's production and sales track record. Clearly, those which capture a greater range of information on the farmer are more useful (and those additional types of data are discussed in the next section on agronomic survey data). Recorded data might only give a partial picture of the farmer cash flow, as some part of the produce might be sold to others and some other crops are produced, but not recorded at all during the sales transaction.

Summary Transaction Data

- Transaction records are generally very reliable and are highly relevant to credit assessment.
- They show evidence of cash flows over time, which help lenders to understand a farmer's capacity to take on debt.
- If already collected for other purposes, agribusinesses could cooperate with lenders to make such data available also for credit assessment at a relatively low cost.

¹⁰ <https://www.provenance.org/tracking-tuna-on-the-blockchain>

Agronomic Survey Data

- More years of experience planting the crop
- A longer track record of selling to the agribusiness buyer
- Above average yields per unit area
- Additional sources of non-farm income
- A history of applying good agricultural practices
- Completed more training on growing a particular crop or managing their finances

Agronomic survey data can be collected by:

- Agribusiness field agents
- Field staff employed by capacity building programs (public-private partners)
- Bank officers or agents
- Mobile phones (based on farmer self-reporting)

Survey data, once digitized, should be of significant interest to lenders. Particularly when administered by trained specialists knowledgeable of the crops under production, such data should be more reliable than if collected via self-reporting or other impersonal data-driven estimates (such as regional statistics, satellite images, etc.).

Collected by Agribusinesses

Some of the most reliable and relevant-to-credit data on small-holder farmers is collected by agribusinesses to manage their cash-crop supply chains. Large agribusinesses are the recognized experts in high-quality cultivation of the cash crops they are buying. They know what farmers need (in terms of inputs and farming practices) to grow crops to the quality standards their consumers demand.

Within the past decade, many agribusinesses have begun to collect large sets of digital data using integrated software platforms designed specifically for the needs of cash crop supply chains (see call out boxes, Farm Force, GeoTraceability-Farm Business Planner below). Most of this data is initially collected by field agents trained by the agribusiness and equipped with mobile devices, and they visit each farm to:

- **Survey the farmers**
- **Measure the farm (with GPS coordinates)**
- **Photograph the farmer and farm**
- **Record baseline data on the farm's current practices for fertilization, weeding, pruning, harvesting, rotation, etc.**

Each farmer is assigned a unique identification number that links its profile information to future activities such as input purchases, pesticide application, and crop sales.

Agronomic survey data resolves many of the challenges lenders face in approaching farmers, including:

- Finding and contacting farmers spread out across wide geographic areas
- Verifying farmer identify and proof of farm ownership
- Estimating the expected production for the farm

Furthermore, a farmer's participation in a structured value chain or contract farming scheme gives lenders:

- A verifiable, third-party source of crop production data
- Certainty of the farmer's access to fair pricing
- Comfort that agronomists oversee the farmer's proper use of inputs to achieve desired yields and minimize crop loss risk

FarmForce

Farmforce was created to help smallholder farmers gain access to formal markets and improve the effectiveness of out grower schemes.¹¹ Its "software as a service" suite combines web-based and android phone interfaces to help exporters manage relationships and operations with a network of smallholder farmers.

A typical FarmForce client will have:

- A network of several hundreds or thousands of growers.
- Field staff that go out to survey the farmers, provide them with training, make sure the right plant varieties are being planted at the right time, manage quality, and provide agronomic advice.
- An agriculture buyer that employs the field staff and manages the export process and relationship with overseas customers.

FarmForce's configurable software offers the agribusiness extensive functionality to collect and monitor the data it needs for many reasons including to:

- Better understand the farmers in their network.
- Manage pesticide residue levels.
- Accurately forecast yields.
- Manage their agent networks/field staff.
- Trace the produce through value chains.
- Comply with quality standards.

The farmer data points FarmForce has found lenders to be most interested in are:

- Ownership document of land (captured by photograph).
- How much land is owned (mapped with GPS mapping tool).
- What is the yield forecast for this year (captured on the platform, but somewhat subjective).
- Crop volumes sold to the company in past years (documented over time).

By analyzing other data on the FarmForce platform, lenders may be able to develop custom credit scores for farmers of a given cash crop.

¹¹ <http://www.farmforce.com/>



GeoTraceability-Farm Business Planner

Geotraceability is a “software as a service” company set up to help integrate independent small-holder farmers into globalized commodity supply chains. Its digitized solutions include data collection tools, traceability systems and online data hosting services that help agribusinesses to collect, process, visualize and analyze data in a cost-effective way, at scale.¹²

One of Geotraceability’s tools, the Farm Business Plan (FBP) Builder, although designed for the needs of agribusinesses, may be of particular interest also to lenders. The FBP Builder helps agribusinesses to consistently and cost-effectively:

- Tailor business plans for each individual farmer in its network
- Communicate with farmers via its monitoring app and/or SMS
- Assess farmer engagement and interventions

The FBP Builder requires the agribusiness farmers to use its field expertise to define agronomic recommendations and set assumptions for the costs of inputs and labour, expected market prices, and seasonal weather forecasts, etc. The FBP Builder then calculates production and revenue forecasts for each farmer.

The FBP Builder should be of interest to lenders because:

- The agribusiness’s agronomists have provided the detailed knowledge of crop economic models that lenders typically lack
- The tool’s revenue forecasts for each farmer are what loan officers have traditionally tried to do manually (and at greater cost) through field visits and interviews.

Potentially, an agribusiness using the FBP Builder tool could approach a financial institution and share its farmer profiles and revenue forecasts. The bank can then screen or score this data to identify farmers it was ready to work with or wished to further evaluate.

¹² <http://geotraceability.com/about/>, accessed 11/18/17

Agronomic survey data is also being collected by other market participants such as:

- Financial institutions (see Yoma Bank call-out box)
- Capacity building programs (see S CPP call-out box)
- Digital service providers (Impact Terra call-out box)

These data sets, which are similar to the data collected by agribusinesses, may also offer a reasonable starting point for the developing of digital credit scoring tools.

YOMA Bank

Yoma Bank in Myanmar has employed two experts from Rabobank of the Netherlands to help develop its agrilending capacity. With the additional support of the USAID Private Sector Development Activity project, the bank has been collecting farmer data to develop a credit scorecard that could further improve the efficiency and consistency of its hire-purchase financing of agricultural equipment. The bank collects data via surveys administered by the agri-equipment dealers’ field sales agents. The questionnaire collects about 60 pieces of data that include qualitative and quantitative information such as:

- Demographics
- Use of certified inputs
- Access to irrigation and post-harvest storage capacity
- Estimated crop and livestock income

Thus far in the trial (December 2017), data has been collected for 1,500 individual loan customers. There have been some issues with the reliability of the data farmers self-report and also with tracking farmer repayments channeled to dealers. The delinquency levels until now have been low, but as the portfolio experiences more delinquent loans over time, it is likely that some of the data collected can be used to develop reliable risk-ranking scores.



SCPP

The Sustainable Cocoa Production Program (SCPP) is a large public-private partnership designed and implemented by Swisscontact, the Swiss Foundation for Technical Cooperation, funded by the Swiss State Secretariat for Economic Affairs (SECO), the Millennium Challenge Account Indonesia (MCA-I) and other big players in the chocolate industry.

SCPP's main objectives are to increase the farmer's household income from cocoa by 75% and reduce greenhouse gas emissions from the cocoa sector by 30%. The program has trained more than 130,000 cocoa farmers in Good Agricultural Practices,¹³ and also aims to improve farmer access to finance, towards which it has trained 50,000 cocoa farmers in financial literacy.

Through CocoaTrace, a cutting-edge software provided by the Indonesian agri-tech company Koltiva, SCPP collects a wide range of data for program purposes that include:

- Transparency/traceability (cocoa companies know where their cocoa comes from)
- Program management
- Reporting to donors (the Swiss and US governments)
- Research¹⁴

Much like the agribusinesses data described earlier, the SCPP program uses trained field agents to visit farms and survey farmers to collect data including:

- Farmer demographics
- Farm size and GPS coordinates
- Products sold by certified farmers
- Farm behavior (or farm practices data)
- Poverty scores (measures of a farmer's likelihood of being below a given-poverty line, and used widely in microfinance for targeting and reporting on outreach goals)
- Data on nutrition and the environment

Currently SCPP is investigating the possibility of linking its farmer-level data with bank data for a smaller subset of its farmers with bank loans. Once enough farmers have taken loans from the bank, and some of those loans have experienced delinquencies, it will be possible to study which data collected by the program best predicts credit risk. A scoring model based on this data can then be used to score SCPP's entire farmer database.

¹³ <http://www.swisscontact.org/en/country/indonesia/resources/library/library-cocoa.html>

¹⁴ Several reports can be found at <http://www.swisscontact.org/en/country/indonesia/resources/library/library-cocoa.html>



Impact Terra

Impact Terra is a social venture developing and implementing digital solutions for small-holder farmers, agribusinesses and other stakeholders involved in agriculture.

Its Golden Paddy platform was launched in Myanmar in December 2016 and now receives over 2.8 million unique visitors each month, covering 97% of townships throughout the country.

Farmers using the app can gain access to the latest and most relevant information about farming practices, weather, pest, and market price information, as well as access to suppliers, buyers and financial service providers. At the same time, Impact Terra is able to collect data on crops under production, farmer demographic and contact data, and patterns of using the mobile app itself. These data on farmers could also potentially be used by lenders to screen farmers for creditworthiness.

Summary Survey Data

- Agronomic survey data is highly relevant to credit assessment.
- If it has already been collected and digitized by others, survey data could potentially be made available to lenders at a reasonable cost.
- When such data has been collected by trained field agents, it is likely to be quite reliable and should complement transaction data well.

Farmer

Demographic Data

Basic demographic data is usually predictive of credit risk. Some well-known examples of 'universal' relationships between demographics and credit risk are:

- **Risk decreases with age, but may again increase at advanced ages (i.e. over 60)**
- **Women are less risky than men**
- **Home owners are less risky than renters**

"Know-Your-Customer" rules require lenders (and many other businesses, such as Mobile Network Operators) to verify the identity of their clients. This is usually done by checking the personal identification of the client. Personal identification information generally contains date of birth, gender and can also include things like nationality/ethnicity.

Other demographic information about household size and location, marital status, and other living arrangements would need to be collected through other means, such as interviews or application forms. Much of the demographic data for farmers is captured by agribusinesses as a part of the agronomic surveys discussed above.

While demographic data can improve the risk-ranking power of credit scoring models, it also effectively discriminates loan applicants on some characteristics over which they have little or no control. In the United States, the The Equal Credit Opportunity Act (ECOA) was enacted in 1974 to prevent discrimination in credit transactions on the basis of race, color, religion, national origin, sex, marital status, and age.

In most worldwide markets, no laws prevent using demographics in credit scoring models and it is often one of the few types of data readily available for credit scoring. However, unless the discrimination is positive (such as favoring women, which is an outreach goal of many international programs and some lenders), it is better to limit use of demographic variables to those that reflect lifestyle (such as marital status, size of household, years in residence, etc.) and avoid those that might be controversial or unfairly disadvantage certain groups (such as race, religion and national origin).

Summary Demographic Data

- Demographic data is usually predictive of credit risk.
- Demographics are one of the few types of data almost always readily available for credit scoring in developing markets financial institutions.
- Some types of demographic data may not be appropriate in certain markets.



Satellite Imagery Data



While farm visits are the surest way to assess land suitability for a given crop, satellite imagery can be compiled on thousands of farms quickly and remotely. Some imagery is available without cost (see call-out box, Sources of Climate Data below) and may include images of the field over many years and planting seasons.

Sources of Climate Data

There are many public domain sources for climate information on the internet. A few prominent examples include:

- The International Research Institution (IRI) Data Library: a repository of hundreds of terabytes of climate-related data. <http://iri.columbia.edu/resources/data-library/>
- The National Weather Service Climate Prediction Center: provides real-time climate monitoring and predictions of climate variability over time scales from a week to seasons, extending into the future as far as technically feasible. <http://www.cpc.ncep.noaa.gov/products/international/africa/africa.shtml>
- Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS): a 30+ year quasi-global rainfall dataset. <http://chg.geog.ucsb.edu/data/chirps/>
- MODIS: vegetation indices, produced on 16-day intervals and at multiple spatial resolutions. <https://modis-land.gsfc.nasa.gov/vi.html>
- Landsat 8: provides high quality visible and infrared images of all landmass as well as near-coastal areas on Earth. <https://earthexplorer.usgs.gov/>

A satellite image of a farm at fine enough resolution (see call-out box on Spatial Resolution below) can validate the planted area and crop type at a site over several years. Specifically, with regard to determining yield from a plot, a Normalized Difference Vegetation Index (NDVI) can be used. This approach seeks to measure the level of photosynthesis occurring in the leaves of the plant, and therefore its likely yield. Descartes Labs, a US firm claims to be able to predict maize yields with a high level of accuracy is a method best suited for annual (rather than tree) crops and to areas with less cloud cover.

Spatial Resolution

The spatial resolution of an image is an indication of the size of a pixel in terms of ground dimensions. It is usually presented as a single value that represents the length of one side of a square. For example, a spatial resolution of 30 meters means that one pixel represents an area 30 meters by 30 meters on the ground. This in turn means that no objects smaller than 30 meters can be distinguished from their background in that pixel.

Horning, N. 2004. Understanding image scale and resolution, Version 1.0. American Museum of Natural History, Center for Biodiversity and Conservation. Available from <http://biodiversityinformatics.amnh.org>.

In addition to adequate resolution, reliably confirming crop suitability requires the farm location to be known very precisely. Currently, much satellite data is likely to be ‘noisy’ or unreliable when attempting to overlay it on very small land plots (i.e. 2 hectares or less) of a small-holder farmer. Some researchers have reported initially promising results using satellite sensor data to predict small-holder productivity in developing markets¹⁵, but for the most part the successful use of satellite data in financial models, such as those used for index-based crop insurance requires:

- Expert knowledge of the particular crop to be insured
- Considerable field work in the locality to understand recent drought episodes
- Rigorous statistical analysis of historic rainfall and vegetation data sets to identify the weather patterns leading up to previous droughts¹⁶

While technology and the internet will continue to lower the cost of obtaining relevant satellite data, making sense of it will still be prohibitively expensive for most lenders, as will adequately understanding crop economic models. This suggests that satellite data would need to be aggregated and analyzed by agribusinesses or software service companies with expertise in data analytics in order to be accessible or affordable to financial institutions.

Weather Forecasts and Records



The analysis of past and current weather patterns to forecast the weather is a special, important and widespread usage of satellite and local sensor data.

The yield of most crops is highly dependent on rainfall and temperature, while humidity can have an impact on crop disease. A farm plot with consistent weather patterns well suited to the crop(s) is more likely to produce higher and more consistent yields.

Hundreds of commercial service providers offer basic and customized location-specific weather forecasting services to businesses. When paired with local agronomic knowledge—such as the type and maturity of the crop under cultivations—weather forecasts can help estimate farm crop yields in the current planting season. Forecasts can also better inform farmers when to plant and fertilize and alert them to any weather anomalies expected during the current agricultural cycle.

¹⁵ Burke, M., & Lobell, D. B. (2017). Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proceedings of the National Academy of Sciences*, 114(9), 2189-2194.

¹⁶ Interview with Dan Osgood, PhD, International Research Institute for Climate and Society, November 1, 2017.

The Weather Channel

The Weather Channel provides both historic weather data and near-term forecasts of rainfall and temperatures.

Historic rainfall and temperature data can establish suitability of a site for the crops a farmer is planning to plant. Seasonal outlook data looks six to seven months ahead and give an indication of any temperature and precipitation anomalies likely within that period. This information can help farmers estimate their yields for the coming harvest. The Weather Channel has been working on a proof-of-concept project providing such data to palm oil farmers in Southeast Asia to help them estimate yields.

While weather is one of the most important contributors to crop yields and crop failure risk in each planting cycle, it is likely to tell lenders only the chances of a given crop’s success and ranges of possible yields in a particular region, as opposed to something specific about a given farmer’s one or two-hectare plot of land. Nevertheless, if interpretable weather data is available to lenders at a reasonable cost, they could benefit from systematically factoring accurate weather forecasting information into their credit assessment models. In addition to complementing transactional data, such lenders can gain an advantage controlling their portfolio quality over a longer-term.

Summary Satellite and Weather

- Satellite imagery can be compiled on thousands of farms quickly and remotely.
- With adequate image resolution and accurate farm coordinates, satellite data can validate planted areas and crop types and estimate yields using a Normalized Difference Vegetation Index (NDVI).
- Weather forecasts can help estimate farm crop yields in the current planting season.
- Satellite and weather data need to be aggregated and analyzed by agribusinesses or software service companies with expertise in data analytics and agronomics.

Credit History



The best predictor of future loan repayment is past loan repayment. That is why so many credit decisions in advanced credit markets rely heavily on credit bureau reports and bureau credit scores. Typical relationships of past credit history to future credit risk include:

- **Timely repayment of past obligations signals lower risk**
- **Spells of delinquency on recent loans indicate higher risk**
- **Frequent requests of credit reports from the bureau indicate higher risk**

For small-holder farmers who have not had credit from a formal credit institution in the past (which is likely to be many or most of them—for example, only 5 % of cocoa farmers in Indonesia have experience with formal loans¹⁷), they would not have a file in a credit bureau. However, developing market credit bureaus also have been steadily improving over time. In Malaysia, for example, bureau coverage has reached 82%, while in Indonesia it has gone from nothing to 18% over the last 10 years¹⁸. Some markets also collect data on digital loan products (for example, Kenya's credit bureau even offers a separate score for digital 'nano' loans). Other markets have dedicated bureaus for microfinance institutions (for example, in Jordan) which may have served farmers in the past.

The key point about credit history is that when third-party objective information about past loan repayment is available, it is likely to be the best predictor of future loan repayment, including for farmers. Credit history most generally dominates any of the alternative data types that will be discussed in the next sections, but those 'new' data types have gained a foothold precisely because so much of the world's population has not had access to formal credit, and alternative data sources offer a new channel to extend credit to them and bring them into the formal financial system.

Summary Credit History

- The best predictor of future loan repayment is past loan repayment.
- Most small-holder farmers will not have a formal credit history.

¹⁷ SCPP (2016): Updated Baselines Report on Access to Finance for Cocoa Farmers in Indonesia.

¹⁸ <https://data.worldbank.org/indicator/IC.CRD.PRVT.ZS?locations=ID>

Alternative Data Sources



Data created through internet and mobile-phone usage are often referred to as 'Alternative Data' or 'Big Data'. These include:

- **Mobile-network operator data (MNO), including use of mobile wallets**
- **Data stored on mobile phones (contacts, calendars, text content, browsing history, installed apps)**
- **Usage data (sometimes called 'metadata'), such as how often people access devices or web-services and how long they spend on various activities**
- **E-commerce data (such as purchase of Amazon, eBay, Alibaba, etc.)**
- **Social media data**
- **Digitally-administered surveys, such as marketing research or psychometric tests.**

The volume and prevalence of such data has grown tremendously in the past decade and use of the internet, mobile phones, and social media is increasingly the rule, rather than the exception, particularly in urban areas.

Most alternative data scoring vendors use data from a combination of these sources. To highlight the properties of the different types of alternative data currently offered in the market, they are presented in three sections:

- Mobile phone/wallet data
- Social media data
- Psychometric testing data

Collecting and mining alternative data requires sophisticated software and specialized skills—for example, some data may need to be ‘scraped’ from devices or internet sites, and pre-processing is needed to ready the data for use in credit scoring models. For lenders, using alternative data is likely to require partnering with data aggregators or vendors selling specialized software and services (at prices that are likely to be negotiated case-by-case).

While alternative data can provide additional, potentially complimentary insight into a farmer’s creditworthiness, such data will come at a cost, so its potential value to credit assessment must be considered alongside cost in each case.

What to Ask Vendors

The recent excitement around big data, machine learning and data science have brought an abundance of start-ups offering proprietary alternative-data credit scoring models. These vendors provide a potentially valuable service, as the aggregation, processing and analysis of large and complex data sets has traditionally not been a core competency of most financial institutions (although this too is changing in advanced markets).

Vendor scores are normally sold to financial institutions on a per score or subscription basis, and the details of the model calculations are not shared. This can make it challenging for lenders to understand which, if any, of similar competing solutions might work best.

Some things to look for and ask about when engaging vendors include:

- Track record of implementation: how long has the firm been in the market?
- References to financial institutions that are successfully implementing their tools
- Evidence from actual use cases: what differences have their models made to business results?

Scoring models are only one piece of a successful digital loan product. Successful credit scoring implementation also (or first and foremost) requires well-designed products, policies and procedures and their competent management, as well as accurate scores and software to implement the scoring model. Lenders are advised that bold marketing claims remain only that until proven (and evidenced) in practice.

Mobile Phone and Wallet Usage

Mobile phone and wallet usage data can provide an objective and verifiable picture of an individual’s regularity and stability of top-up behavior, call patterns, and, in the case of mobile wallets, cash inflows and outflows.

Mobile data usage patterns related to lower repayment risk include:

- Longer registration on the mobile network
- Clients who top up their phones by larger amounts (and less frequently)
- Larger spends on network airtime

The main challenge in using mobile network data for credit scoring is obtaining the data itself, both for model development and on an ongoing basis. The availability of Mobile Network Operator (MNO) data will differ by market. In Kenya, leading mobile operator Safaricom sells anonymized views of its data to financial institutions.¹⁹ In other markets, obtaining MNO data could require specific negotiations or partnership arrangement with a given MNO.

Increasingly, vendors are also offering solutions for calculating credit scores based on mobile phone/wallet usage. These may be based on data they obtain in partnership with a MNO or based on an app the potential borrower chooses to install in order to be able to provide lenders with a credit score (see call-out boxes T-CASH Indonesia, and WeCash and Tunai Kita below). Such vendor technology is promising in that it gives lenders an insight into the credit risk of applicants who may not have any traditional credit history and would otherwise be difficult for the banks to reach or affordably assess. The challenge for lenders working with such vendors is in managing the risks of lending based on models they do not see or understand (sometimes referred to as ‘black-box’ models).

T-CASH Indonesia

Indonesia mobile network operator Telkomsel is launching a digital loan product for its T-Cash mobile money users. Users will be able to apply for 30-day loans ranging from 500,000 Rupiah to 2 million rupiah (approximately \$35 to \$150 USD) without collateral.

T-Cash loans and limits will be offered based on a credit-scoring algorithm that favors subscribers with a longer track record and stable mobile network usage patterns. Interest rates will also be set based on network behavior.

The mobile money loan targets micro users outside of major cities and will be offered in cooperation with several banks including Bank BTPN, BNI and Mandiri. The product launch data is yet to be announced.

<http://www.cnnindonesia.com/teknologi/20171206113246-213-260446/t-cash-nanti-bisa-beri-pinjaman-tanpa-agunan>

¹⁹ Safaricom Kenya reports not the actual usage data for individual clients, but in which ‘band’ of values (by decile) the client’s actual data is located.

WeCash and Tunai Kita

WeCash is a Beijing-based credit scoring platform that asks users in China to register with a cellphone number and to connect to them via social media and Taobao (a Chinese online shopping website) accounts. The user then receives a credit score, from A to D, that can be shared with interested parties, which could include lenders, P2P lending sites, renters, or potential employers.

With joint-venture partners PT Kresna Usaha Kreatif and PT JAS Kapita, WeCash set up PT. Digital Tunai Kita (DTK) in Indonesia. Using technology analogous to WeCash's China platform, the company currently offers its credit assessment services in the Jabodetabek area of Indonesia.

Although the primary focus of both companies has been unsecured consumer finance, both WeCash and DTK are actively investigating ways to serve the agriculture market. In China, WeCash is working with the International Finance Corporation and banks to pilot WePig, a data model that evaluates creditworthiness of hog farmers. DTK is currently investigating the possibility of providing credit evaluation tools for financing rice farmers in Indonesia.

<http://wecashgroup.com/>
<https://blog.tunaikita.com>

There are several 'success stories' of digital credit scoring used for large-scale consumer lending. The biggest and best of those remains the M-Shwari savings and loan product in Kenya²⁰. Its digital scoring model looks at mobile phone and money usage patterns to estimate the size of loan (if any) a borrower is eligible for. Its loans are for very small amounts for a period of 30 days (the most common loan size is \$2 USD equivalent, with a median of \$10 and average size of \$30).

Further research is needed to understand how strongly and consistently a farmer's personal mobile phone and money usage is related to willingness and ability to repay a longer-term loan for an agribusiness. Trends around mobile phone usage that predict well for consumer credit risk may not necessarily carry over well to agri-loans. Mobile wallet purchases and receipts will be excellent for prediction if related to the farm business (and thus covered in the Transaction Record section above), but personal receipts and purchases maybe be less so.

Other potential barriers to the effectiveness of mobile network data for farmer credit assessment include:

- Limited connectivity (and low mobile phone usage) in some rural communities
- The cost (or possibility) of lender access to mobile network data

²⁰ <https://www.cgap.org/sites/default/files/Forum-How-M-Shwari-Works-Apr-2015.pdf>

Social Media and Networks

Some fintech startups are using social media data and mobile-device data (such as number of contacts, social media usage patterns, content of SMS messages or even regularity of charging the phone) to verify identity and/or rank loan applicants by risk of repayment for the purpose of consumer lending.

Some examples of social media characteristics related to lower credit risk include:

- A connection with someone who has successfully repaid one or more loans²¹
- Use of more sophisticated language in posts
- Having an active profile in LinkedIn (a network for professionals)

Social media data should also be able to help confirm that a person is who they claim to be. This could be of value in situations where national identification numbers are difficult to verify.

While network relationships and communication behavior patterns can certainly highlight some differences between people, this data is generally too big and costly for lenders to process. Like weather data, it would require a data aggregator or analytics company to process it in a way that can be consumed easily by lenders at an affordable cost. Lenders would need to see its value in ranking its clients by risk compared with, or in complement to, other risk assessment tools.

For the time being, it is likely that social media data is less relevant to small-holder farmers. While some farmers in some parts of the world may be increasingly active on social media, many are limited by constraints such as:

- Lack of consistent network coverage in rural areas.
- Lack of mobile phone ownership.
- Relatively homogenous small-holder farmer demographics—for example, in parts of Asia, a large share are men in their 50s.

²¹ Jeevan Vasagar, Financial Times, JANUARY 19, 2016, Kreditech: A credit check by social media, <https://www.ft.com/content/12dc4cda-ae59-11e5-b955-1a1d298b6250>

Psychometrics

Psychometric testing has the potential to extract insights on any loan applicant's personality. A well-designed psychometric test might measure an applicant's 'character' in the same way a loan officer does through 'reference checks' with business partners and neighbors (and without the need to inform other people about the loan application).

Personality traits that have been related to lower repayment risk in studies include²²:

- Higher "Conscientiousness"
- Higher "Integrity"
- Higher "Cognitive Ability"

Unlike traditional lending character checks, which are based on personal observation and independent references, psychometric testing relies on applicants answering questions designed to measure particular personality traits.

While psychometric testing has a long and successful track record of usage in human resources and clinical psychology, its accuracy measuring personality traits in lending situations is largely unstudied and unknown. Several vendors, however, claimed success using psychometric tests to rank loan applicants by risk of loan repayment.

Entrepreneurial Finance Lab

Entrepreneurial Finance Lab (EFL), the first commercial company to develop credit assessment based on psychometrics, now has over 10 years of experience applying psychometrics to loan repayment. By July of 2017, 1 million people had taken its tests for credit assessment.²³

EFL is currently working with the specialized ag-finance company Juhudi Kilimo²⁴ and MasterCard Foundation to pilot their credit assessment models on small-holder farmers in Kenya.

Since May of 2016, over 7,000 tests have been completed, with over 4,000 loans issued to the test-takers (although loan decisions were based on Juhudi Kilimo's standard procedures, rather than on the tests). While there are still few delinquencies overall, farmer acceptance of the tests has been good, and the preliminary risk-ranking results have met expectations.

EFL is also testing out new 'chatbot'-based tests that can be administered over feature phones using SMS messages, which offer the potential to reach and pre-screen farmers in remote regions.

In November 2017, EFL merged with Lenddo, which uses mobile & digital footprint data for credit scoring. For more information on LenddoEFL, see www.include1billion.com

²² Caire, Dean, <https://www.business-school.ed.ac.uk/crc/wp-content/uploads/sites/55/2017/02/What-Personality-Measures-Could-Predict-Credit-Repayment-Behavior-Dean-Caire-Galina-Andreeva-and-Wendy-Johnson.pdf>

²³ <https://www.eflglobal.com/about/>

²⁴ Juhudi Kilimo, Transforming the lives of farmers, <http://juhudikilimo.com/>

Although personality tests can rank loan applicants by risk and have the tremendous advantage of being potentially available from any applicant, the method requires considerable cost and expertise to develop. There is also a time and money cost for lenders who wish to administer them²⁵ and the method is subject to several challenges including:

- People understand and respond to questions differently based on their education and backgrounds.
- People answer questions differently in different situations.
- Questions about personal preferences and values have not traditionally been asked of loan applicants.

Summary Alternative Data Sources

- Collecting and mining alternative data requires specialized software and skills.
- Most lenders will need to partner with mobile operators, data aggregators or vendors in order to use alternative data for scoring
- Alternative data insights into farmer creditworthiness should be considered alongside cost and competing options in each case.

²⁵ Such psychometric tests require 20 to 30 minutes to administer and vendors generally charge per test

Putting it All Together – Which Data Sets Are Complementary?

Each type of data discussed above has some cost to collect and, for any given value chain, some types of data are likely to work better for credit scoring than others.

The mechanics of building credit scorecards are not discussed in this guide, but are explained rather simply, yet in some detail, in the IFC's recently published Data Analytics And Digital Financial Services handbook.²⁶

Once a few predictive borrower characteristics or behaviors (risk indicators) are combined in a credit scoring model, each additional 'risk indicator' contributes incrementally less to the model's overall risk ranking power. This modelling reality favors use of the data that is:

- Lowest cost
- Most objective and reliable
- Most relevant (intuitively) to ability/willingness to repay a loan
- Can be most consistently collected from all applicants/borrowers

This means that when a financial institution has enough data, it should give preference to data points that:

- Are objective and can be observed directly, rather than reported by the applicant
- Evidence relationships to credit risk that confirm expert judgment
- Cost less to collect
- Can be collected from most, if not all, applicants
- Do not discriminate based on factors the borrower cannot control (i.e., age, gender, race) or that are subject to prejudice (i.e., religion, ethnicity, language)²⁷

²⁶ See section 1.2.3 Analytics and Applications: Credit Scoring, beginning on pg. 79 at <https://www.ifc.org/wps/wcm/connect/22ca3a7a-4ee6-444a-858e-374d88354d97/IFC+Data+Analytics+and+Digital+Financial+Services+Handbook.pdf?MOD=AJPERES>, accessed 11/17/2017

²⁷ Also taken from the IFC Data Handbook section written by the same author

Chart 1 provides a visual summary of the data sets reviewed here along the dimensions of data relevance, availability, cost, reliability and expected predictive power in relation to loans to farmers. It is not based on empirical evidence from particular sets of data, but on the author's own judgment and personal experience working with similar types of data.

Chart 1: Data Set Ranking Table

Data Set Ranking Table	Relevance	Availability	Cost to Lender	Reliability	Predictive Power
Credit History 	HIGH	LOW	LOW	HIGH	HIGH
Transaction Records 	HIGH	HIGH	LOW	HIGH	HIGH
Agronomic Surveys 	HIGH	HIGH	LOW	AVERAGE	AVERAGE
Demographics 	AVERAGE	HIGH	LOW	AVERAGE	AVERAGE
Mobile Phone/Wallet 	AVERAGE	AVERAGE/LOW	AVERAGE/HIGH	HIGH	HIGH
Psychometrics 	AVERAGE/LOW	HIGH	HIGH	LOW	AVERAGE
Social Media 	LOW	AVERAGE/LOW	HIGH	LOW	LOW
Satellite 	HIGH	AVERAGE/LOW	AVERAGE/HIGH	AVERAGE	LOW

Ideally, a balanced scoring model would contain elements of at least credit history, transaction records, agronomic survey data and lifestyle-related demographics (marital status, household size, years in address, etc.). Such a model might be further augmented by alternative data sources if they were consistently available at a reasonable cost. However, it is possible to approach the market from different angles, and first-movers might find that various combinations of these data types are effective enough to rank farmers and facilitate credit decisions for a particular use case.

The Value of Digital Scoring to Agribusinesses

The potential of digital credit scoring presents an opportunity for companies who buy from farmers as well as the agro-input companies who supply farmers with fertiliser, seeds/seedlings and crop protection products.

At present, many farmers use informal loans from family, fellow farmers or local businessmen. The World Bank reports that informal lending tends to be short-term, which precludes longer-term investments, and thus only partially meets farmers' financial needs, usually at a high cost²⁸. If the costs and limits of informal credit leads farmers to use a minimum of inputs, the result can often be relatively low yields.

Low yields present two key challenges for buyers:

- They face higher transactions costs from dealing with a larger number of lower yielding farms.
- Low yields constrain supply, which puts upward pressure on prices.

For an agribusiness, a 10% increase in productivity from its existing farmer network is far preferable to increasing its number of network farmers by 10% (given the fixed costs of working with additional each farmer).

For lenders, the agribusiness's agronomic expertise, reliable farmer data, and access to markets should greatly reduce the barriers (real and perceived) to working with farmers. Agribusiness data could lower the costs of client acquisition and due diligence and provide reliable, verifiable information about farm production. In addition, the involvement of the agribusiness to some extent ensures the farm is working to quality standards.

The next sections look at some ways agribusinesses and lenders might collaborate to turn this digital data into credit assessment tools and financing that can benefit lenders, agribusinesses and the small-holder farmers. However, as the call-out box, The Limits of Technology below reminds us, lending trials or new credit products should always keep in mind the everyday realities and needs of the farmers in rural communities.

²⁸ World Bank, Focus on Sustainability 2004, Our Business, <http://siteresources.worldbank.org/ESSDNETWORK/Resources/481106-1129303936381/1777397-1129303967165/chapter2.html>



The Limits of Technology

The digitization of farmer data is already facilitating progress in agricultural production and food security. It seems that digital data could do the same for farmer access to finance.

At the same time, it is important to remember that technology alone cannot solve all problems of financing small-holder farmers. A different line of thinking with a long tradition (and perhaps less familiar to the 'digitization' crowd) is that farmers may still, in the digital age, prefer the flexibility of the informal finance they have worked with for years in their small, rural communities²⁹.

²⁹ See Rick Van Der Kamp's blog <https://www.linkedin.com/pulse/six-myths-farmer-finance-rick-van-der-kamp/>, accessed 11/15/2017 or, for a more extensive account of similar arguments, the paper: Adams, Dale W; and Robert C. Vogel. (1986) "Rural Financial Markets in Low-Income Countries: Recent Controversies and Lessons", World Development, Vol. 14, No. 4, pp. 477-487.

How to Engage with Financial Institutions

For lenders not specialized in agriculture, lending to small-holder farmers with traditional underwriting processes has often not seemed to make business sense. Loan values are small, the distances to the farms can be great, and the associated risks (of crop disease, weather events, and market access) are perceived to be high. Furthermore, their loan officers and staff generally lack expertise in farming.

Agribusinesses working in structured cash crop value chains, on the other hand, already have data that meets many of the lender's traditional underwriting challenges. Its agents visit and review the farmer, provide valuable skills training and assure market access for the produce and collected the data lenders need to assess credit worthiness. It seems like an obvious 'win-win', yet realizing potential synergies requires taking a first step, and agribusinesses may be best positioned to do this.

Where to Start?

Lenders are (rightly) likely to be hesitant to rush into a market long perceived as risky, no matter what kind of data or vendor model they are presented with. Cooperating with established agribusinesses seems like an obvious way for them to access relevant data at a reasonable cost and test the concept of digital scoring as a tool to lend to small-holder farmers at scale.

As with any innovation, it is necessary to prove the business case for lending to farmers based on agribusiness data. A proactive agribusiness could approach a lender (or possibly a mobile network operator) with plan to 'pilot test' data-driven lending using a phased approach such as:

- 1 Use agribusiness data to pre-screen some subset of farmers (i.e. with above average yields or a longer track record of sales) for further due diligence by the lender (as in the FarmForce example mentioned in Page 19). Pre-screening is a very simple, rule-driven variation of scoring. Additional farmer data may also be collected in this stage through questionnaires or working with data aggregators/vendors.

- 2 Design a standardized and relatively short-term, small loan product and issue a cohort of loans to the selected farmers.
- 3 Use the repayment performance data on this first cohort of loans to develop a basic statistical credit scoring model using the agribusiness data.
- 4 Develop larger, longer-term and/or more nuanced loan products for 'graduated' farmers who have successfully taken and repaid the basic loan product.

The key principals of a phased-approach, in addition to someone taking the first step, are the gradual extension of credit, starting with small, simple and short-term products. Farmer positive credit history unlocks larger, longer-term credit. Such a strategy, which has been successfully used for digital consumer loans, may also work well for farmers.³⁰

Other Possibilities for Developing a Scorecard

Lenders can also look to involve other partners to pilot lending approaches to farmers, or even for the express purpose of developing a digital scorecard. Many initiatives aimed at digitization of small-holder farmer data are already supported by donors such as the MasterCard Foundation. Possible approaches for dedicated programs to test digital scoring for farmer loans include:

- Bringing together multiple parties that have mutual interests in piloting a structured loan product. The call-out box, Trial 1: Digital Loans for Inputs below provides an example of a mobile network operator, an input provider, and an agribusiness working together to provide a relatively small loan for inputs over a digital channel.
- Working with a lender that has already made loans to some subset of those farmers in the past (i.e. if some farmers have any personal or farm-experience using formal financial institutions). As discussed in the call-out box, Trial 2: Capacity Building Program Partners with a Lender below, a bank can possibly analyze farmer data alongside its own past lending experience to develop a statistical scoring model and began lending to the lower risk farmers.
- Make some number of loans (i.e. 1,000 loans) to small-holder farmers for which the agribusiness has collected data without any credit assessment and treat the expected losses as a business development cost (or, possibly, attract funding for a 'guarantee' to offset losses in the interest of developing a better scoring tool for small-holder farmers in a large value chain). If enough of these first 1,000 loans go 'bad', or experience serious delinquency, the data could be adequate to develop a scoring model free of the 'selection bias' usually present when lenders develop models by analyzing past data only for clients they deemed safe enough to give loans to. It is important, however, in the context of such a pilot, to go on to collect repayment of delinquent loans, otherwise the market could be ruined by signals that loans from a given financial institution need not be repaid.

³⁰ This example does not include satellite and weather information on the assumption that the agribusiness has already ensured farmers are planting appropriate crops and no adverse (uninsured) weather events are expected for the season.

Trial 1: Digital Loans for Inputs

The fintech arm of a mobile network operator (MNO), a fertilizer producer and a coffee buyer are currently in discussion with GrowAsia to conduct a lending trial in Southeast Asia. The interests of each of the parties in the trial are as follows:

- The fintech is trying to promote its mobile money product and expand its credit portfolio. The partnership allows the company to explore a closed loop model which reduces both acquisition costs and risks
- Grow Asia is generating learning on the value of lender, input provider and buyers' partnerships
- The fertilizer producer is looking to increase sales of its fertilizers the market
- The coffee buyer hopes to significantly expand its coffee purchasing in the region
- The pilot will allow it to explore how it can leverage its farmer data and access to farmers to improve input usage, and therefore yields

One of the key issues that come up in multi-party trials such as this are data privacy and data sharing among the parties. In many countries, the fintech will need to ask the farmer for consent to use the data collected about it by the agribusiness.

The other thing to remember about trials such as this is that they must involve enough farmers that some critical mass of farmers experience repayment problems—there are no strict numeric rules, but any fewer than 100 delinquent loans is likely to leave too much of the results to chance—namely, the chance that the farmers in the trial are not representative of all the farmers in the target population.



Trial 2: A Capacity Building Program Partners with a Lender

Indonesia's Sustainable Cocoa Production Program (SCPP) program has studied data on its 116,000 farmers and found that 5.3% have formal loan experience and a credit report, while 1.9% have an outstanding loan.

Given the extensive data SCPP collects, it is in discussion with lenders to see if it can match its data to the credit history of its farmers who have past borrower experience. If there are enough loans, and, particularly, enough loans with past delinquency, SCPP and/or the lender may be able to use the combine SCPP's program data and credit history data to develop a statistical scoring model for the cocoa value chain. Even if this will be possible, it will need to be used cautiously, perhaps by first applying it to farmers with a bank account but no prior borrowing. A model developed using data from farmers with a bank account and applied to farmers without a bank account is less likely to work as expected, as those two types of farmers may be significantly different in terms of their characteristics and behaviors.

No matter what approach is used to move towards digital scoring, this guide advocates a principal of transparency, gradual engagement and shared risk-taking to allow small-holder farmers to build credit histories. Digital data and its possible use for credit scoring does not mean lenders should go from zero to 60 miles per hour in entering the small-holder farmer segment. Realistically, the longstanding challenge of formally financing small-holder farmers profitably and sustainably is unlikely to be solved in a day by a fintech start-up with a magic wand (or even a very good prediction model).

Finally, whether digital data is used initially only for screening or it is possible to immediately develop some sort of digital scoring model the loan product itself and processes for administering it must be well-designed, appropriate to farmer needs, and well-executed. Some considerations for product design are discussed below.

The Importance of Loan Product Design

A well-designed loan product, including related policies and processes, is in fact more important to success than any scoring model that is used for decision making.

Agri-loan products should match the disbursement and repayment of loans to the needs of the crop season. Farmers periodic cash flows may be limited, and the repayment schedule should reflect this.

Mitigation of some other types of repayment risks, such as potential misuse of loan proceeds, can be creatively designed into agri-loan products and processes (such as paying the loan proceeds to input suppliers and collecting the loan repayments from agriculture buyers), while mitigating the greatest risks, such as crop failure due to extreme weather or disease events, may require affordable insurance schemes or guarantee funds to offset.

Regardless of the risk ranking power of any credit scorecard, it alone will not make a successful loan product. Although an adequate discussion of product design is beyond the scope of this digital scoring guide, product design's importance to a successful loan product, and even to the results of trial intended to develop a credit scoring model, is crucial.

A well-designed product, and associated loan disbursement and collection processes, should help to limit problem loans due to any reasons besides ability and willingness to repay.

Two rather straight-forward ways to control operational risks around the disbursement and collection of loans are to pay loan proceeds directly to the input supplier and to collect loan repayment directly from the ultimate crop buyer. Each of these is discussed briefly in turn in the next section.

Input linked Loans

Lenders can structure digital loan products so that the loan proceeds can only be used for the purpose of purchasing inputs. This is important, because it is possible for a farmer to ask for a loan for improved inputs, but instead use the loan proceeds to pay for school fees, medicines, or other purposes. If the premise for the loan was to finance inputs that should improve yields and revenues (and make loan repayment possible), other uses of the loan funds might make timely repayment difficult. However, building positive credit history through input-linked loans might also help that same farmer take out a consumer loan for school fees in the future.

Buyer repayment deduction

Much as in the case of disbursements, collecting loan repayment directly from buyers ensures that the farmer repays the loans in the first instance—in order to receive money for the crop. This approach is likely to work only in well-structured value chains. For a well-designed loan product, particularly one operated on a digital platform or using mobile money for repayment, such an arrangement could make loan repayment easier for the farmer, collection easier for the lender, and potentially lower the associated fees.

Risks and Challenges

The great potential of digital data to facilitate credit assessment of small-holder farmers brings with it the great responsibility of ensuring the well-being of those same farmers and their local communities. Some of the risks and challenges lenders need to consider in lending to farmers at scale over digital channels include:

- Protecting farmers' rights to data privacy in light of relevant laws in each country. Farmers should give explicit consent to use their data for loan evaluation.

Understanding the strengths and limitations of the scoring model. A farmer's bureau score (if available) will always be relevant to credit risk on any new loan, but the predictive accuracy of other types of data need to be tested with the specific loan products and target farmer population to be known.

- Being sure the digital loan products and their terms and conditions are understood by farmers. Poor communication about loan conditions or issues with the platforms that process and administer the loans can lead to repayment problems that were not measured during scorecard development.
- Protecting farmers against over-indebtedness. Scoring algorithms rank borrowers by risk, but credit policy and business rules related to loan size and term should be informed with agronomic know-how and appropriate to cashflow cycles of farm production.

Conclusions and Recommendations

This guide has presented some types of farmer and farm-relevant data currently collected and digitized by agribusinesses and other service providers. It has also suggested that the first efforts to develop digital data scoring tools can address the estimated 40% of small-holder farmers in structured cash crop value chains. The data on these farmers is likely to be better, and the farmers themselves more creditworthy. Still, a phased approach, with the gradual extension of credit, starting with small, relatively simple and short-term products, is recommended. It will integrate the farmers into traditional systems of data collection used by lenders while providing lenders with repayment performance 'feedback' that is needed to inform more sophisticated scoring models. At the same time, the approach purposefully establishes farmer credit history as the key to unlocking larger, longer-term credit.

Some additional recommendations (to each of agribusinesses, lenders, and partnerships) on enabling a data environment conducive to digital scoring are:

Agribusiness

Consider making electronic payments to farmers

The use of mobile wallets (and other electronic payments) offer a range of advantages from traceability to reduced theft. Digital payments provide the farmer with a verifiable cash flow record that can help accessing loans in the future, and digital payments are easier than cash for lenders to collect.

Leverage the data you already have

If an agribusiness records a farmer's sales, training attendance, and demographics and/or location, make this data available to financial institutions. It could help them to secure lower cost credit.

Lenders

Use agribusinesses data

Most agribusinesses want their suppliers to succeed and grow. Leverage any data they can share on farmers transaction history, demographics, training attendance, etc. to reduce or corroborate other due diligence efforts. Agribusiness data can potentially lower client acquisition costs and improve the accuracy/reliability of risk measurement.

Leverage agribusinesses knowledge

Agribusinesses have extensive knowledge of the crops they buy. Lenders can lean on this expertise to develop and sharpen their own competency in agronomic analysis, commodity price projection, and input selection.

Links loans to productive inputs

Supply loans specifically to purchase productive inputs such as a fertiliser, quality seed or new machinery. Most agribusinesses have an informed view about which inputs will improve yields and cash flows.

Start small and stage growth

Start with small loans and treat the delinquent loans as a model development cost. It is important to extend enough loans to be able to get repayment performance data and improve credit risk models. Once processes for scoring, distribution and repayment are proven with small loans, larger loans can be offered to farmers who have successfully repaid smaller loans.

For Partnerships

Openness

The agribusiness and finance industry have different languages, business metrics and decision timeframes. Take the time to establish the partnership and clearly document the roles and objectives of each partner.

Innovation

Scoring with digital data to assess the credit risk of small-holder farmers in structured value chains is of great interest to many, but there are few examples of success in practice. Instead, successful use cases of credit scoring for small-holder farmers have tended to rely on data collected in labor-intensive field visits by loan officers and/or the specialized agronomic knowledge in certain microfinance institutions. Since developing digital data scoring models for farmers will require some trial and error, partnerships can also benefit from involvement of development agencies that may have budget to test new methods—not all of which will succeed.

In conclusion, the longstanding challenge of formally financing small-holder farmers profitably and sustainably is not a task for artificial intelligence or machine learning—it is a task for agribusinesses, financial institutions and stakeholders in the international economic development community. It is the collection and intelligent analysis of digital data, the cooperation of stakeholders in trials of agri-loan products, and a focus not only on corporate profit, but on improved food security and livelihoods for small-holder farmers worldwide, that can come together now to make digital credit scoring and affordable formal credit for small-holder farmers a reality in the not-distant future.

Acknowledgements

Grow Asia and SAFIRA thank **Dean Caire** for his efforts in undertaking the background research, interviews and authoring the report. We would also like to thank the following people for taking the time to speak to Dean in phone interviews and provide comments to draft versions of the guide. Their insights are incorporated and reflected not only in the cases specifically referencing their companies, but also informed the overall content of the guide.

Hans Dellien, International Finance Corporation

Bailey Klinger, Entrepreneurial Finance Lab

Dirk Lebe, PhD, Sustainable Cocoa Production Program, Indonesia

Sarah Giselle L. Lopez-Pozas, Nestle

Spencer Morely, FarmForce

Dan Osgood, PhD, International Research Institute for Climate and Society

Roy Parizat, World Bank

Steven Ripley, Geotraceability

Bas Rozmuller, International Finance Corporation

Michael Wang, The Weather Channel



SAFIRA

Strengthening Agricultural
Finance in Rural Areas



Australian Government

The project is part of AIP-Rural, a partnership between Indonesian
and Australian Government for rural economic development.