Insight2 Impact

Advancing Financial Inclusion

# Exploring the spatial data landscape and options for sustainable data collection

February 2018

Established by

cenfri

FMT FINMARK TRUST

Technical Note

## Author

David Taylor, Data4FI, insight2impact

## About insight2impact

Insight2impact | i2i is a resource centre that aims to catalyse the provision and use of data by private and public-sector actors to improve financial inclusion through evidence-based, data-driven policies and client-centric product design.

i2i is funded by the Bill & Melinda Gates Foundation in partnership with The MasterCard Foundation.

### For more information:

Visit our website at **www.i2ifacility.org.**
Email Mari-Lise du Preez (i2i's Partnerships Manager) at **mari-lise@i2ifacility.org.**
Call us on **+27 21 913 9510.**

# Contents

# Introduction

This paper explores the current landscape of spatial data for financial inclusion, with a focus on sustainable collection and management. It is important that it be considered in context and a reflection of the current state of the broad range of approaches and activities that are taking place within markets that aim to improve financial inclusion through the increased and improved use of spatial data for decision-making.

This is a complex topic with no silver bullet and no one-size-fits-all solution. In a bid to recognise this complexity and provide a functional framework for discussion, this paper divides into three sections the issues that surround sustainable spatial data collection. Section 1 examines various data collection frameworks. It looks at the strategic and high-level question of who should, or who is, driving change in the data market. Section 2 examines data collection methodologies. It also examines the approaches and options that have been tested and that are being implemented in various markets, and it examines their suitability in various contexts. The final section gives an overview of the tools available for collecting spatial data and their relative advantages and appropriateness for different markets.

# Spatial data collection frameworks

**This section examines a number of data collection frameworks, which can be thought of as the high-level strategic approaches to data collection. A useful way to categorise these frameworks is to understand why the data is being collected and which institution requires that the data be collected.**

These frameworks can be viewed as standalone options, with some being appropriate in some markets, some being appropriate in others and with hybrids and combinations being appropriate for most. They can also be viewed as a journey, with the appropriateness of each being judged against the current priority and change that is desired in the market.

For example, if you are looking to stimulate interest in a market with little understanding of, or exposure to, the potential uses of spatial data, this may require an approach that depends heavily on outside support (e.g. donors). While this approach may not be sustainable in itself, it may be a necessary step to generate the conditions in public and private actors to allow a sustainable approach to be developed. The appropriateness of different solutions should be evaluated regularly in light of changes and evolutions of market conditions.

It should also be noted that these options are not mutually exclusive and an appropriate solution may be complex and may be a combination of one or more approaches. The purpose of this section is not to dictate an exhaustive list of options to be picked from but to provide the framework to facilitate discussion and explore the complexity of the approaches that are available.

Each option will be evaluated against the criteria in Table 1 below.

### Table 1: Evaluation criteria

| Format | Latitude |
|---|---|
| Complexity | The logistical, bureaucratic, organisational, social or regulatory complexity that would have to be overcome in order to implement the framework |
| Sustainability | The chances that the framework will produce timely, high-quality information that meets the needs of financial service providers (FSPs), regulators and other public organisations, on an ongoing basis |
| Catalyst | The ability of the framework to generate insights that will catalyse the market to move towards a more complete and integrated use of data to drive financial inclusion |
| Market information | The ability of the framework to produce data and information that are likely to increase understanding of the market and the financial inclusion landscape |
| Data availability | How readily the framework lends itself to the development and maintenance of a common platform to share and analyse spatial data |
| Ownership | How appropriate the framework is to support or incentivise regulators, or other policy-making bodies, to provide an institutional home for a platform to house a public, consolidated, spatial inventory of financial access points |
| Institutionalisation – public | How appropriate the framework is to support or incentivise regulators to make use of the data to make decisions that positively affect financial inclusion, i.e. by setting aspiring but realistic financial inclusion goals |
| Institutionalisation – private | How appropriate the framework is to support or incentivise FSPs to make use of the data to make decisions that positively affect financial inclusion, i.e. by using the data to target and improve service in unserved or underserved areas |

The criteria will be measured on the following scale: high – medium – low – none.

# Donor/market facilitator-driven data collection

Kenya, Nigeria, Tanzania, Uganda, Zambia, Bangladesh and India have all conducted field-based mapping of financial services, to some degree. Some of these projects have been undertaken in partnership with regulators or central banks, but they have been primarily driven by a market facilitator, such as a Financial Sector Deepening Trust (FSD) and funded using grants from donors, primarily the Bill & Melinda Gates Foundation[1]. This situation was, in turn, driven by the recognition of regulators and bodies such as the Alliance for Financial Inclusion (AFI)[2] of the importance of being able to measure proximity (accessibility of users to financial access points) as a component of financial inclusion.

## Evaluation

| | |
|---|---|
| Complexity | Low |
| Sustainability | Low |
| Catalyst | High |
| Market information | High |
| Data availability | Medium |
| Ownership | Low |
| Institutionalisation – public | Medium |
| Institutionalisation – private | Low |

## Advantages

This approach has been incredibly valuable in stimulating interest in, and demand for, spatial data in the financial inclusion space. It is unlikely that many markets would have advanced to their current state without this approach being adopted. The approach has created real data and allowed discussions about the possible value of spatial data in improving the understanding of financial inclusion to move from the abstract to the concrete. It has allowed the development of use cases and has institutionalised the importance of measuring proximity, as a component of financial inclusion, within financial inclusion measurement frameworks. In its simplest form, proximity is measured by using proxies, such as the number of functional service points per 10,000 population of a given administrative area. This can evolve into a more direct measure of proximity (like that adopted by the Bank of Tanzania, which measures the percentage of the population within five kilometres of a financial access point) and is likely to evolve further into more nuanced measures that establish proximity targets based on service type (at the financial access point) and demographic, economic and geographic profiles. An example of this would be setting a target of having 100% of the population of urban centres be within one kilometre of a functional mobile money agent but recognising that, for example, a target of 20% of the rural population within 10 kilometres of a bank branch may be more realistic. A donor/market facilitator-led approach provides the initial baseline data for regulators to explore use cases, develop measurement frameworks, set appropriate targets and build up institutional buy-in before investing heavily in new policy.

## Challenges

Following this approach involves a number of challenges. Firstly, it is unsustainable. Donors and market facilitators cannot, and should not, be relied upon to fund expensive and time-consuming,

1 https://www.gatesfoundation.org/What-We-Do/Global-Growth-and-Opportunity/Financial-Services-for-the-Poor
2 http://www.afi-global.org/members/

field-based data collection surveys indefinitely. While this approach may, in many cases, be a necessary one to start the journey, it does not solve the important and long-term issues of institutionalisation and ownership that are key to driving usage and real change.

Secondly, this approach has not directly resulted in broad, in-depth use of spatial data within financial service providers. There are many reasons for this, including inadequate human and technical resources to make use of the data, but there are also barriers that have been created by the data collection process itself. These include:

- **Access** – In many markets the spatial data is published only after a long delay; and, in some cases, it is not made publicly available at all. In other cases, the data is only made available through gatekeepers who impose an application process before the data can be accessed. While these may sound like trivial limitations, they can and do have a marked impact on eventual uptake and usage of the data.
- **Anonymised and aggregated data** – In a number of markets, the spatial data is only made available in aggregated form, meaning that financial institutions cannot differentiate their own service points from anyone else's. This limits the usefulness of the data to financial service providers to the point where it is almost impossible to use it for strategic or operational decision-making.
- **Lack of standardised unique identifiers** – The existing datasets are missing standardised unique identifiers for the points collected. While traditional financial service providers, such as banks, could link their internal and transactional data to the location data by using the branch names, even this can be a tedious and time-consuming process. For digital financial service providers (DFSPs) with tens or hundreds of thousands of access points, it becomes a practical impossibility to link the locational data to their internal records. Again, this trivial-sounding problem imperils the whole idea of sustainability by removing the ability of the service provider to mainstream location data into their analysis and decision-making processes.

## Characteristics of markets where donor/market facilitator-led data collection may be appropriate and/or more likely

- Where a regulator, or central bank, has interest but does not fully appreciate the potential of spatial data to enhance measurement or delivery of financial inclusion
- Where no publicly available spatial data currently exists
- Where there is understanding and commitment from a regulator to use the data but financial constraints exist that prevent them from collecting the data
- Where there is no movement from the private sector to collect spatial data, independently of regulation
- Where it is felt that proximity is a key factor influencing the uptake and usage of financial services
- Where expansion of digital financial services has reached such a level that self-reporting of existing agent locations would be an undue and unfair burden to service providers

# Regulator-driven spatial data collection

Regulators play a key role in driving financial inclusion through the setting of policy and by creating an enabling environment. Regulators set policy and targets for financial inclusion and are a key consumer of spatial data to drive more informed decisions. They can also play a key role in stimulating the market to make use of relevant data through mandating which data is collected and reported by FSPs. The increasing recognition of the value of spatial data for regulators can be seen from this quote from the AFI Financial Inclusion Data Working Group:

*"The rewards are valuable enough to regulators that all 42 members of FIDWG are currently in various stages of developing financial inclusion GIS maps. This stands in contrast to two years ago when only a handful of countries (Nigeria, Tanzania, Uganda, Kenya, India, Bangladesh and Indonesia) had geo-spatial mapping projects, most of which were conducted as a part of the Bill & Melinda Gates Foundation's Financial Services for the Poor mapping project."*[3]

In its most explicit formulation, regulators can set policy to require FSPs to report accurate spatial data for their financial access points. An example of this is Tanzania's E-Money Regulations 2015, which require electronic money service providers to submit agents' and their physical location identity, full addresses (including physical address) and GPS coordinates[4]. A number of countries have made attempts to encourage FSPs to report the locations of their services, including space for GPS coordinates in their reporting templates. However, these are rarely backed up with an explicit policy that compels FSPs to report these and, even where there is such a policy, it is rarely followed up and strictly enforced.

The inclusion of spatial data in an increasing number of reporting templates is a positive move and shows the increasing recognition of the importance of this data for proper market oversight. However, a lack of high-quality, standardised reporting templates and inadequate ability to enforce the reporting requirements often hamper the implementation of a well-intentioned initiative.

## Evaluation

| | |
|---|---|
| Complexity | High |
| Sustainability | High |
| Catalyst | Medium |
| Market information | Medium |
| Data availability | Medium |
| Ownership | High |
| Institutionalisation – public | High |
| Institutionalisation – private | Medium |

## Advantages

One of the major advantages of driving change through the regulator is that, once a regulator has mandated that spatial data be collected and reported for financial access points, financial service providers are compelled to collect this data, whether they have developed use cases and the potential value for their own organisations or not. By compelling FSPs to self-report, not only can the

3 http://www.afi-global.org/blog/2015/06/geo-spatial-mapping-financial-inclusion-has-proven-its-worth
4 Bank of Tanzania: E-Money Regulations 2015 (Part IX, 39.1)

regulator ensure the sustainable collection of data but (by mainstreaming the collection into FSPs' workflow and data systems) it is also removing significant institutional barriers within FSPs which could otherwise prevent them from collecting and making use of the data themselves.

Driving the collection and reporting of data through regulatory requirement also ensures that the data have an appropriate institutional home and ensures that the necessary information be available for the regulator to appropriately guide the market to implement solutions that will support increased access to and usage of financial services.

A regulator-led framework also makes it more likely that spatial data be collected using common standards and that quality control be undertaken on the data. Even in markets where it is not appropriate for a regulator to mandate self-reporting and collection of spatial data, the regulator still has an important role to play in developing and publishing standards and guidelines to ensure that if FSPs collect data on a voluntary basis it can be aggregated and analysed in a meaningful way. Where there are no commonly recognised standards, it is conceivable that a situation could arise where FSPs independently decide to collect their own spatial data using different terminologies, coding systems and coordinate systems. While these issues could theoretically be overcome to allow the data to be aggregated, it would likely be a complex and time-consuming task for the regulator and would likely complicate the development of any future standardised reporting mechanism.

## Challenges

One of the largest challenges with regulator-mandated self-reporting in markets with developed digital financial services (DFS) landscapes is to ensure compliance by DFS providers. Telecoms companies, banks or fintechs who have spent a significant amount of time and money expanding the geographic reach of their services are likely to feel aggrieved by being compelled to undertake an expensive and time-consuming exercise to map their existing services. This will be especially true of markets where DFS service providers have adopted the "aggregator" or "super-agent" model, whereby no actual staff of the service provider would visit an agent service point in the normal course of their work. This would further increase the cost and time it would take to collect the data, as it would be much more difficult to mainstream the data collection into the normal activities of the super-agent or aggregator than it would be with service provider staff.

Another challenge for this approach is the time and effort that policy change can take to implement. Reporting requirements and policies are often reviewed infrequently, and it could easily take years for the opportunity to come around to influence policy if the regulator is not highly motivated to drive through the requirements. Even where regulators are fully committed to driving change by compelling reporting, in some countries consultative processes and procedures can seriously delay the enacting of the policy or reporting requirements.

It is often the case that responsibility for regulating the financial services channels is split between many different regulators. This can lead to a situation where many regulators have to be brought together to agree and implement meaningful reporting standards. While some countries, like Nigeria, have successfully used coordinating groups and steering committees to define common reporting needs goals that have been used to develop common reporting standards, it should be recognised that there is also a risk that the bureaucratic complexity that can arise in coordinating multiple institutions can significantly delay the process if the effort is not accompanied by a strong and committed political will guiding the process.

### Characteristics of markets where regulator-led data collection may be appropriate and/or more likely

- A strong regulator that can develop and enforce spatial data reporting requirements (which is key for this framework to be successful)
- Markets that have a relatively small number of DFS points already in existence (in which it is more likely to gain support from FSPs)
- A regulator who wants to incorporate an understanding of proximity and coverage into its financial inclusion metrics
- Markets where the regulator has produced, or is in the process of producing, formal reporting requirements for the establishment or registration of digital financial service points

## Data collection driven by financial service providers

There is growing recognition of the power of data-driven decision-making within financial service providers. While there are still significant barriers for many institutions to fully realise their aims, there is broad momentum in the market towards an increased use of data. For many institutions, spatial data is seen as a key ingredient of a fully data-driven institution.

Although there are likely to be many reasons for this recognition and the institutional changes that it is driving, for our purposes here it may be useful to think of this move as being intrinsically linked to increased competition in the market and the rise of disruptive products and services. While many studies are being carried out to better understand the role that data currently plays in the decision-making processes of FSPs, one trend is already becoming apparent. Younger companies focused on providing digital financial services (such as telcos and fintech companies) are much more likely to be data-driven than older financial institutions with far greater institutional inertia (such as commercial banks). However, as competition increases and older institutions reposition themselves to enter the digital economy, they are also being forced to undergo more fundamental organisational changes to make more and better use of data. This presents a valuable opportunity for market facilitators to influence both types of institution to ensure that the institutions have the incentives, capacity and data to ensure that these changes affect financial inclusion positively.

Several DFSPs, in markets where there is no regulatory requirement to do so, have already invested in collecting spatial data. This ranges from institutions who collect the coordinates of their service points on using hand-held GPS units as part of their registration process (which are then often not used or are underutilised) to institutions who not only routinely collect spatial data on their own services but who also collect spatial data on competitor services, other relevant businesses and service providers and use this data to drive operational and strategic decisions – from calculating and projecting (for instance) the cost of fuel for the distribution of marketing material to identifying potential new markets for products and services.

## Evaluation

| | |
|---|---|
| Complexity | High |
| Sustainability | High |
| Catalyst | Medium |
| Market information | Low |
| Data availability | Low |
| Ownership | Low |
| Institutionalisation – public | Medium |
| Institutionalisation – private | High |

## Advantages

The biggest advantage of this model, and its variants, is that it not only provides a sustainable solution that produces regularly updated, high-quality data for measurement by the regulator but equally, if not more importantly, it is also an important indicator of data usage within FSPs. A data collection approach driven by the FSPs indicates that the market has recognised the value of data in general and spatial data in particular and that they have invested in processes and systems to capture and make use of the information. This is a crucial step in the journey to get markets and institutions to use data to improve financial inclusion. If an institution or market has already committed to using the data for their own priority use cases, then half of the job of the market facilitator is already done. What remains is to demonstrate the commercial value of spatial data in a way that positively affects financial inclusion and to facilitate filling technical gaps, such as capacity issues and availability of contextual and supplementary data in the market. With new data products and tools (such as the WorldPop suite of demographic and poverty maps) being developed to help providers identify commercially viable unserved or underserved markets, the case for using data to improve financial inclusion and private-sector revenue is growing.

Another distinct advantage to FSPs driving change and opting to collect and make use of spatial data without regulation is that it does not require the intense effort and delays that are often associated with changes in policy by the regulator.

## Challenges

One challenge that can arise with this approach is that it can produce non-standardised datasets. While this may sound like a trivial detail, small issues (like a lack of standardised coding, the use of different projection or coordinate systems for spatial data or even something as simple as differences in spelling of a provider name) can cause major problems and exponentially increase

the effort and resources required by a regulator or central bank to combine the data, conduct meaningful analysis and develop, or measure, financial inclusion metrics.

Another challenge is that it is not only possible but probable that not all private-sector players will commit the resources necessary to collect spatial data at the same time. This leads to a situation where it may be possible to get an accurate and detailed picture of some sections of the market but one that is incomplete and so does not provide the complete picture that a regulator or central bank may need.

While a key driver for institutions who have started collecting their own spatial data is using it to inform their customers and potential customers of the locations of their products and services (an exercise that would entail making the data available to the public), there is nothing inherent in this approach that would ensure that the data be made available publicly or to the regulator. This approach would still require a third party, which could be a regulator or other actor, to aggregate the data and host and manage a common platform that allows market actors to derive insights from it.

## Characteristics of markets where FSP-led data collection may be appropriate and/or more likely

This approach is more likely to grow organically, where:

- Private-sector software companies develop appropriate data collection and analytic tools and then market these directly to financial service providers
- At least one significant service provider is capturing and making use of spatial data (Competition can be a strong motivator; and once a large FSP is known to be exploiting data that provides a competitive advantage, it is more likely that others will follow suit.)
- The capture of spatial data is imported from one country to another by a service provider as standard or best practice
- Digital financial services are still in their infancy, or the market is relatively small, and the effort required to collect spatial data for existing service points is relatively low

> “FSP-led data collection is more likely to grow organically where at least one significant service provider is capturing and making use of spatial data.”

# Data collection driven by the private sector

An evolution of the donor/market facilitator framework would be for a private company to conduct census mapping of all financial access points and potentially several other point types that might inform decision-making for multiple public-sector and private-sector actors in multiple markets, i.e. agriculture, businesses and health services. The difference would be that this data would not be published publicly. It would be held as the intellectual property of the company that collects the data and would be sold to financial service providers and or regulators who wish to make use of it. While this approach has not been tested in the market, there has been enough interest in exploring the approach theoretically that it is worth mentioning.

## Evaluation

| | |
|---|---|
| Complexity | Low |
| Sustainability | Medium |
| Catalyst | High |
| Market information | High |
| Data availability | Medium |
| Ownership | None |
| Institutionalisation – public | Low |
| Institutionalisation – private | Medium |

## Advantages

The biggest advantage of this approach is that it would be quick, would likely produce high-quality data and would be sustainable – if demand for the data were sufficiently high. This approach would not require the long delays that a regulator-led approach might entail and would avoid the risk of partial data collection that could be the result of an FSP-driven approach. If we accept that this data is valuable for decision-making, then speed is surely of the essence in markets that are continually being transformed by disruptive products and services.

Another advantage of this approach would be that it could easily be used to collect and maintain spatial data from a wide range of other sectors, such as agriculture or health. This data could be used to inform the development of sector-specific products, such as agricultural insurance or health-focused savings products or to identify and target unserved or underserved populations and so increase both access and drive usage.

## Challenges

The primary challenge to this approach is the significant risk that would be taken by the private company that would have to invest in the technology and person time that it would take to collect and keep the data current. This risk could be hedged or mitigated altogether by a market facilitator or a consortium of public and private stakeholders who could agree to license the data for a fee before the field teams collect the data.

Another challenge associated with this approach is that, although the data is technically available to the market, licensing costs introduce a barrier that may prevent uptake and usage by FSPs and other actors, which could significantly inhibit its ability to drive improvements in financial inclusion.

This approach could possibly also limit the ability of FSPs to use the data to inform their own customers about the locations of their products and services. It is likely that this issue could be addressed in the licence agreement, but it is one that should be recognised.

### Characteristics of markets where private-sector-led data collection may be appropriate and/or more likely

- Smaller markets with a relatively small number of financial access points (This is likely to be a market with a small DFS market. This is important as the larger the market, the greater the cost of data collection and so the greater the risk for the company that collects the data.)
- Where there is a need for spatial data to be collected for other sectors (This could lower both the cost and the risk for the data collection company.)
- Where there is a strong and publicly expressed need for the data from actors, either in the public sector or the private sector, willing to pay for access to the dataset but who are unable to collect it themselves

> “Private-sector-led data collection may be appropriate and/or more likely in smaller markets with a relatively small number of financial access points, where there is a need for spatial data to be collected for other sectors.”

5 Bank of Tanzania: E-Money Regulations 2015 (Part IX, 39.1)

# Case studies and hybrid models

**It is unlikely that any of these approaches in their rawest form will provide a complete, wholly appropriate, solution in any given market. Below are two examples of how the frameworks above can be combined to meet the individual and unique needs of diverse markets.**

## Facilitated registration and self-reporting hybrid in Tanzania

Tanzania has fully integrated the use of spatial data into its financial inclusion strategy, to the extent that one of its key measures of financial inclusion is the percentage of the population within five kilometres of a financial access point. Tanzania also requires all electronic money issuers to report the GPS coordinates of their access points.

*"An electronic money issuer shall maintain and submit to the bank records of recruited agents that shall include: (a) agents' and their outlets identity, full addresses, including physical address, GPS coordinates"*[5]

Although this requirement is not currently actively enforced, these pieces have merged into the design of a regulator-owned registry of financial access points. While it is recognised that self-reporting will have to provide the backbone to any sustainable data collection model, there are some interesting and unique conditions in Tanzania that have led to the development of a hybrid model that takes elements from three of the frameworks mentioned above.

Tanzania has a well-developed DFS market with the number of agent points now numbering in the hundreds of thousands. It has also moved to a model of agent recruitment and management, which has meant a distancing of the service provider and the agent. The clear majority of mobile money agents in the market are now recruited and managed by aggregators or super-agents. While this is good news for financial inclusion, it provides a significant barrier to FSP-led self-reporting. Because of the large number of agents and the fact that no staff directly employed by the service provider have direct contact with agents on the ground as part of their normal work streams, it would be incredibly costly for service providers to enter the market and collect the spatial data for all the tens or hundreds of thousands of existing agent points. It would also entail significant duplication of effort, as a single agent point could provide services on behalf of all multiple service providers and a simple self-reporting model would require all networks to visit that point to collect the data individually. Finally, simple self-reporting would disproportionally disadvantage FSPs who are having the largest impact on improving geographic access to services. The more service points an FSP has and the more remote they are, the more it will cost to collect and report this data. This means that a simple, regulator-mandated self-reporting framework (while possible) could have negative and lasting effects on the willingness and ability of FSPs to engage with other aspects of improving financial inclusion.

Because of this, an approach has been developed which would leverage the previous experience, and funds, of the Financial Sector Deepening Trust Tanzania (FSDT) in partnership with the Bank of Tanzania (BoT) to carry out a field-based exercise to register all existing financial access points in the country, which would include issuing a BoT ID sticker to all locations that provide financial services. This would remove the financial barriers and inequalities that would be produced by a simple regulator-mandated self-reporting model, which would require FSPs to map all existing agent locations.

In parallel with the mass registration by FSDT/BoT, the tool used to register existing points will be made available to all FSPs. FSPs will be required to mainstream the registration of new locations and capture the spatial data as part of their existing onboarding processes. This will include empowering FSPs to register new locations and issue IDs on behalf of BoT in locations that are not already providing financial services.

This approach cherry-picks the advantages from a range of other approaches to create a mechanism that not only provides a sustainable data-collection mechanism but also mainstreams the data collection and so, hopefully, data use within FSPs. This approach will create a sustainable registry of financial service points, which will meet the data needed to measure and improve financial inclusion and will open up new possibilities for consumer protection products and services, by allowing all financial services to be linked to a verified location and by allowing information on poor service levels, fraud and other complaints to be shared not only between service providers and the regulator but also across and between FSPs.

## Creating champions and appropriate environments for FSP collection

Another hybrid approach, which is being discussed in some markets where there is no regulation to compel FSPs to collect and report the locations of their service points, is for market facilitators to identify market champions (FSPs) who are committed to using spatial data to identify opportunities and underserved markets but who need support to do so. The market facilitator then works with the market champions to diagnose the factors that are preventing them from taking full advantage of spatial data in their decision-making.

These issues could be related to:

- Human capacity in the organisation
- Human capacity available in the market
- Organisational information systems
- Organisational decision-making systems
- Availability of contextual information for decision-making, or
- A variety of other constraints

Once the diagnostic exercise has identified the barriers, the market facilitator works with the FSPs and other relevant actors who produce, supply, manage and analyse data to address the key barriers to usage.

Meanwhile, the market facilitator works with the regulator to define data-collection standards and best practices. This could range from developing Excel-based data-reporting templates to developing standardised data collection tools, which are explored in the next section.

By increasing the institutional value of the data to FSPs and lowering the barriers and cost of mainstreaming the collection through the provision of standards and tools for data collection, this approach aims to catalyse a market where FSPs are incentivised through competition and market pressure to self-report in a sustainable and standardised way that allows them to make the journey at their own pace while incentivising those who are ready to adopt early through the provision of additional technical support.

> “Once the diagnostic exercise has identified the barriers, the market facilitator works with the FSPs and other relevant actors who produce, supply, manage and analyse data to address the key barriers to usage.”

# Data collection models

**This section examines the different models that can be used to collect spatial data. It focuses on questions like "who collects the data" and "how is it collected".**

The first three models focus mainly on the collection of spatial data for existing financial service locations, although there is a case to be made that crowdsourcing could provide a sustainable solution to data collection, and the last focuses on mainstreaming the capture of location data into existing workflows.

## Field-based census mapping by a private data collection company

### Methodology

Kenya, Nigeria, Tanzania, Uganda, Zambia, Bangladesh and India have all conducted field-based census mapping of financial services to some degree. This methodology involves paid enumerators walking the streets to identify financial service points and then using a smartphone-based app to record their location and conduct a short interview to collect data on the services provided.

While in most markets this kind of exercise has been done in partnership with the regulator and service providers, in certain markets it has been carried out with a large degree of independence from them in order to produce the data quickly to develop and showcase use cases and to use it to stimulate the market for such insights.

### Advantages

This approach provides, complete, high-quality data for all service providers in a relatively short period. It also allows for the collection of location data for other services that may be of interest to financial service providers and regulators, including health services, schools and agricultural supply chain infrastructure, such as markets and agricultural goods suppliers.

This approach also has the advantage, from the view of a donor or development partner, that it is relatively easy to fund and that it provides a methodology that is easy to replicate across a wide range of countries with very different markets contexts. Indeed, a single company, BrandWorx, has conducted census mapping of financial access points in all of the countries listed above. Because of the level of standardisation provided by this approach, it has been relatively easy to develop a common platform to house and analyse this data for multiple countries[6].

### Challenges

While this approach has gone a long way to stimulate and drive forward the use of spatial data to increase financial inclusion, the relatively high and centralised cost of data collection means that it is not a suitable solution for sustainable, long-term, regular data collection. Donors and development partners are not willing or able to fund these massive data collection exercises indefinitely, and there are no signs that regulators or FSPs would be willing to fund these exercises on a regular basis.

As well as being unsustainable, there are some other challenges with this approach. Firstly, due

6 The first version of this tool can be found at http://www.fspmaps.com/

to the cost of this kind of census, it can only be carried out infrequently (perhaps every other year, at most). While this may be adequate to measure financial inclusion in markets with a high penetration or rapid growth of digital financial services, the data goes out of date incredibly quickly and so is not an accurate reflection of the market and, therefore, has limited use to drive policy or investment decisions.

Secondly, it is never possible to capture all the financial service points with 100% accuracy. Mobile money agents and savings and credit cooperative organisations (SACCOs) can be challenging, as they can operate in places that may not be visible or accessible to data collectors (such as office buildings and multi-storey shopping complexes), or they may not be clearly signed and visible from the street. There is also the problem of mobile money agents who have no fixed place of business but rather carry their till with them and provide services for friends, co-workers or family. These issues can lead to an underestimation of the access points that are available – especially in high-density urban areas.

A methodological challenge that has arisen in most markets where this approach has been taken is the difficulty of linking the spatial data that is collected with the transactional data for that service point held by the service provider. This is a particular problem for digital financial service points and stems from the lack of a systematic unique identifier for agent location. For example, if a mobile money agent is recorded in the census as "Juma Mohammed", firstly, there may be several hundred or thousands of agents with the same name. Secondly, any variation of spelling or order of the names will make it impossible to match the spatial data captured in the census to the transactional and agent records of the service provider. In some markets, till codes have been used to try to solve this problem, but this has caused other problems when the tills move about, when, in interoperable markets, agents cannot only provide services for multiple providers but can also have multiple tills for the same provider and the fact that it is not easy for enumerators to accurately record strings of numbers without making mistakes.

Because this model does not lend itself easily to regular cost-effective updating, it proves difficult to institutionalise the data that is produced once an institution has taken ownership of the data and the responsibility to keep it current. They are also responsible for funding further rounds of expensive data collection indefinitely if there is no plan for a more cost-effective data collection methodology.

## Appropriate Markets

This kind of mapping has been used in several countries to demonstrate the value of the insights that spatial data can produce. It has provided the concrete data for events like FinDisrupt[7] to use to excite the market in Tanzania and to stimulate the creation of spatial data working groups, as we have seen in Uganda. It is useful for markets where no spatial data relating to financial service points currently exists, as it allows conversations to move from the abstract into the concrete, and it allows stakeholders to get hands-on with the data and start generating and understanding the insights that can be produced.

7 http://www.fsdt.or.tz/findisrupt/findisrupt-1/

This kind of approach can be particularly useful as a starting point in markets with high penetration of digital financial services. This is especially true where the relationship between the service provider and the agent is managed through an intermediary such as a super-agent or agent aggregator. In these kinds of markets (where there may be hundreds of thousands of service points and no direct contact between the service provider and the service location), there can be significant resistance from DFSPs to allocate the resources necessary to conduct the census mapping of their existing agent network, which would require a mass data collection exercise by staff who would not ordinarily be in direct contact with the agents.

The challenges of census-style self-reporting are complicated in these markets by the fact that the cost of this exercise would be disproportionately higher for service providers whose digital financial services are more disaggregated, not concentrated in urban centres and that are likely to be serving the mostly rural population. An initial census-mapping approach undertaken by a third party for all existing financial service points can help level the playing field and reduce the barriers to the introduction of a more sustainable mechanism for data collection, such as mainstreaming the collection of spatial data into FSP onboarding processes.

This can be contrasted to other markets with far fewer digital financial service points and where the service provider directly recruits and manages the agents. In this kind of market, there may be only a few thousand or low tens of thousands of agents and staff, directly employed by the service provider, that have regular, direct contact with the agent or service point in order to:

- Conduct training
- Provide support and promotional material
- Collect information
- Distribute airtime or other products
- Perform other supervisory tasks

In these cases, very little additional cost would be involved if the service provider were required to capture the location data for all of its existing agent networks, as part of its existing activities.

> “An initial census-mapping approach undertaken by a third party for all existing financial service points can help level the playing field and reduce the barriers to the introduction of a more sustainable mechanism for data collection.”

# Community/ crowdsourcing

## Methodology

Crowdsourcing encompasses a variety of slightly different data-collection methodologies, but what they all have in common is that they rely on individuals or organisations to contribute data through a common mechanism. At its most extreme, this could mean that individual community members are asked to use their own smartphones to report the locations of financial services, although this has never been seriously proposed. In the context of mapping financial service points, crowdsourcing is most likely to involve an existing group with a common interest, coming together to map the locations of financial services. The Bill & Melinda Gates Foundation has conducted two pilot studies to test this approach, using Humanitarian OpenStreetMap (HOSM) (a collective of students who volunteer their time to support spatial data collection) in Uganda and Premise (a private company that relies on community members who are remunerated for conducting surveys) in Nigeria[8].

This approach can be thought of as an iteration of the field-based census mapping by a private company. The methodology is essentially the same, the difference being that one methodology relies on employees and the other relies on a self-selected group or community. The major hypotheses tested in the pilots were that this methodology could produce data of comparable, or higher, quality to the data collected by a private company at a price point, which would make it cost-effective enough to be funded on an ongoing basis by regulators, market facilitators or the financial service providers.

## Advantages

The organisations who managed the data collection in the Bill & Melinda Gates Foundation-funded pilots, HOSM and Premise, were confident that their methodology could be used to produce data of a similar quality to the census mapping by a private data collection company at a lower price point.

## Challenges

This methodology suffers from many of the challenges that the first census approach suffered from. While it may be true that it could lower the cost of data collection, it is unlikely that it will lower the cost dramatically enough to provide a long-term, sustainable solution. It is, further, unlikely that regulators or FSPs would be willing to fund this kind of data collection on an ongoing basis.

Some crowdsourcing models rely on local, community-based data collectors to report data to a central location. These models could provide a solution to more timely data collection, as community members could report new services as they appear. However, this model would require a vast number of highly motivated enumerators spread all across a country for it to produce regularly updated, high-quality data. It is unlikely that this approach would work at scale.

Crowdsourcing methodologies also require a significant level of trust in the individuals or organisations collecting the data. There could be perceived problems with legitimacy of the data collected, and regulators may not be comfortable

8 A full description of these pilot projects can be found in Building sustainable geospatial data resources for financial inclusion October 2016 by Loeb and Mutemi at http://www.i2ifacility.org/Publications/Building%20sustainable%20geospatial%20data%20resources%20for%20financial%20inclusion.pdf

entrusting the collection of data to measure important financial inclusion metrics to volunteer students.

This methodology also does little to engage financial service providers or encourage them to mainstream data collection or use spatial data to improve decision-making and so improve financial inclusion. It provides an alternative to the first methodology as a possible mechanism to collect data to help measure access to financial services, but it is insufficient to drive the kind of institutional change within FSPs that is needed to increase the availability and appropriateness of financial services.

### Appropriate Markets

This methodology may be a viable and cost-effective alternative to the private company census mapping in small markets with highly active mapping communities or private crowd-based data collection companies. It could be used to produce a baseline dataset that could be used to stimulate interest and engage the market, but as it faces many of the same challenges as the first methodology it is unlikely to provide a long-term, sustainable solution on its own.

## Census mapping of existing financial service locations by FSPs

### Methodology

This method requires FSPs to capture and report the locations of all their existing service points.

### Advantages

One of the main advantages of this methodology is that it removes the requirement for outside funding. In markets with a small number of financial service locations, where FSP staff regularly visit all their service points, this methodology could be effectively used to collect a complete baseline dataset without significant investment or inconvenience to the FSPs.

Another main advantage is that it institutionalises the data-collection process within the FSP, and this could lead to increased usage of the data to generate insights that could positively affect financial inclusion.

### Challenges

The challenges to the implementation of this methodology depend on several things, including:

- The total number of existing service points in the market
- The highest number of service points of any single FSP
- The availability of spatial data for existing points within FSPs
- The physical size of the country and the distribution of service points
- Whether agents of FSPs are regularly visited by staff during their normal activities

In markets where FSPs have not always collected spatial data on their service points, the larger the market and the more disbursed the service points are, the costlier the data collection will be and the greater the resistance from FSPs will be. This will be amplified further if DFS agents are not regularly visited by FSP staff during their normal activities. The high cost of these activities may well be passed on to consumers in the form of higher prices, which would ultimately not serve the aims of increasing financial inclusion.

Even in markets where there is a regulatory requirement that location information for all service points be reported to the regulator, it has proven extremely difficult to enforce.

### Appropriate Markets

This methodology may be appropriate in markets with small DFS networks where FSP staff have physical contact with their service locations as part of their normal activities. Under these circumstances, the level of effort and investment required from the FSPs, to capture and report the spatial data, would not be overly burdensome. In this situation, the data quality and standardisation will be dramatically improved if a central body, ideally a regulator, sets clear and explicit reporting standards or, even better, develops a standardised data capture tool for FSPs to use to collect and report spatial data during their normal activities.

## Mainstreaming of spatial data capture during recruitment, registration or onboarding of new financial service points

### Methodology

This method focuses on the capture of spatial data for new service points, as opposed to the previous two methods, which focus on capturing the data for already existing points.

Any truly sustainable data collection methodology (which hopes to produce regular, high-quality data on the location of financial services in a way that will encourage its usage by both FSPs and regulators to improve financial inclusion) will, eventually, have to include mainstreaming the capture of spatial data as part of the recruitment, registration or onboarding processes.

The requirement to capture spatial data at the recruitment phase of a new service point could be imposed by a regulator or could be mandated from within an FSP.

### Advantages

Mainstreaming the requirement to report spatial data for new service locations would require minimal investment from FSPs, as agents or staff are required to physically visit the site of the service during the recruitment or onboarding process. It also ensures that the spatial data is available to FSPs to meet their priority use cases and is linked to their own systems and identifiers,

allowing them to easily link up their locational and transactional data.

If linked to a central data repository, this methodology would provide real-time location data for all service points opened after its establishment.

## Challenges

Many FSPs still use paper-based forms to record the information for new service locations during recruitment. These then require digitising before they can be used for analysis. This digitisation process involves transcribing data from paper into a digital format, and even with modern Optical Character Recognition software the error rates are often too high to produce accurate spatial data files. This challenge could, however, be seen as an opportunity, as transcription issues affect all data collected on a new service point, not just spatial data. This is driving a move towards digital capture of information in the field. It is likely that FSPs, especially those who provide services through an agent model, are likely to switch from paper-based data capture to app-based registration technology (regtech), which would not only provide a higher quality of basic text data but would also allow for easy capture of spatial data, images and potential biometric data for identification.

This methodology does not solve the issue of capturing data for existing service locations.

## Appropriate Markets

It is easy to see how this method provides a solution to producing almost-real-time, high-quality, spatial data for financial service providers; and if properly combined with another methodology for capturing the data for existing service points, it can form the backbone of a sustainable market information system.

“Many FSPs still use paper-based forms to record the information for new service locations during recruitment. These then require digitising before they can be used for analysis.”

# Data collection tools

**This section will provide an overview of the kinds of data collection tools that are available to capture spatial data. It is likely that this section will go out of date as new technology becomes available. It should be read as a review of the options available in early 2017.**

## Smartphone apps

### Methodology

There is an increasing number of smartphone-based data collection apps, which allow you to conduct in-depth surveys and collect a wide range of data types, including:

- Single and multiple-choice questions
- Free text, integers or decimals
- Location data
- Pictures, video and audio input
- Scans of barcodes or QR codes

Many of these apps are built on freely available, open-source libraries, such as OpenDataKit (ODK) and OpenMapKit (OMK), which is itself an extension of ODK.

Surveys are designed and uploaded to a server, users download new surveys to their smartphone apps, as needed. These surveys are used to record location and supplementary data on the phone.

### Advantages

When used correctly, apps allow the collection of high-quality, structured data. Apps can be used to validate data entry in the field, to check that the right kind of data is being collected in the correct format. For example, a simple validation check would ensure that a phone number be entered only using numbers and that the phone number be the correct length.

Apps also allow offline data collection. Data can be stored locally on the phone or on a removable memory card in areas where a reliable data connection is not available. The data can then be sent manually or automatically when the phone is in range of a data connection.

Furthermore, apps are relatively easy and cheap to deploy and there is a growing number of service providers who provide data-collection software – a service arrangement where they will provide technical support, manage the data hosting and even help design the forms in return for a monthly subscription fee. Most apps also offer the option of hosting the database onsite within an institution, although this option is usually significantly more expensive and can be less flexible.

### Challenges

Data-collection apps run on smartphones which need to be paid for. If this data is collected as part of a dedicated collection effort by a research company, financial services provider or other dedicated data collection institution, then – most of the time – the devices are provided by the company that manages the data collection process. This often entails locking the devices down, to ensure

that only a limited range of functions be accessible to users. This is done to ensure that important settings are not changed, that batteries are not drained and that enumerators are not using the phones for their personal communication or entertainment needs. This provides the control required to ensure the quality of the data but comes at a cost, which can be a barrier for institutions to implement or mainstream data collection into their normal processes.

While data collection models that rely on some version of crowdsourcing often make use of the personal smartphones of the data collectors (which is obviously cheaper than buying dedicated phones for data collection), this often means that there is an increased need for training and ongoing technical support to ensure that the tools are working in the field. With suitable smartphones becoming progressively cheaper in many markets, the price difference between providing smartphones and the increased cost of training and support narrows significantly.

### Most useful for:

Large-scale data collection, where users who are familiar with smartphones use the app to collect data or register financial access points

## Bots

### Methodology

An internet bot – also known as a web robot, www robot or just as a bot – is ”a software application that runs automated tasks over the internet”[9]. For our purposes, we are interested in “chat bots”, which we can think of as programmes that mimic human conversation and that can be used to collect data from users in a more conversational and organic fashion, which can be done using smartphone apps.

Chat bots allow simple programmes to be written for messaging services like WhatsApp, Facebook Messenger, Viber and Skype. These programmes can then ask users questions and collect data using pre-programmed responses. Bots can do the work of a call centre full of humans at a fraction of the cost. i2i is currently exploring the possibility of creating a data-collection chat bot to allow users to test and explore the technology.

### Advantages

One of the biggest benefits of using bots to collect data from members of the public is that it does not require them to download an additional app or software. The clear majority of smartphone users will use a messaging service such as WhatsApp, Facebook Messenger, Viber, Skype or WeChat, which has either launched or is developing application programming interfaces (APIs) to allow outside developers to launch bots using their messaging service.

9 https://en.wikipedia.org/wiki/Internet_bot

No or very little training is required to use the bots, as instructions can be built into the data-collection conversation.

Bots are an effective way to deliver and collect information. Their ability to provide the user with the experience of holding a conversation with a real person builds a relationship that can be used to deliver information and drive behaviour change.

Bots can be used by dedicated field data collectors or by members of the public. Their ease of use makes them an incredibly flexible way to collect data.

Data collection can either be initiated by a human, or questions or surveys can be initiated by the bot at scheduled intervals.

Bots can be programmed to function in multiple languages and can collect a range of data types, including location data, pictures, video and audio recordings.

## Challenges

The use of bots for data collection and communication depends on people's access to smartphones and a working data connection.

Bots are very new and offer a range of exciting options for the future of how we think about data collection and communication. However, as they are new, they have not been as extensively field-tested as data collection apps, and more conservative institutions who want to fund and implement tried-and-tested methods may see this as a disadvantage.

## Most useful for:

Collecting structured data from non-dedicated data collectors, i.e. financial service users, where it would be difficult or impractical to provide training.

> "Bots are an effective way to deliver and collect information. Their ability to provide the user with the experience of holding a conversation with a real person builds a relationship that can be used to deliver information and drive behaviour change."

# Web scraping

## Methodology

Web scraping involves searching the internet, usually the websites of financial service providers, for details of the locations of their services. This "scraping" of the website can yield the following possible results:

- **No information:** A considerable number of financial service providers have no information on their website about the location of their services. This is especially true of digital financial service providers, such as telecoms companies. Some financial service providers, especially savings groups and non-formal providers, may not have a website or any online presence at all.
- **Addresses of service locations:** Service providers often publish the street addresses of their major service locations. This is especially common for commercial banks. While these addresses can sometimes be turned into data that can be mapped, a process known as geocoding often relies on very poor-quality data and results in information which cannot be used.
- **Coordinate-based data:** An increasing number of financial service providers are starting to invest in using technology to help customers and potential customers to find their services. This has led to more service providers adding online maps to their websites. It is often possible to extract the GPS coordinates of these services from the maps. However, care should be taken with this kind of data, as many service providers simply drop a pin on the map where they think their services are located, instead of mapping them in the field. This approach can lead to a large variation in the quality of the data, with some institutions mapping to the building level, others to the street level and others simply putting the pin in roughly the right location.

## Advantages

Theoretically, web scraping can provide a level of spatial data without having to go into the field to collect locations.

Web scraping can be used to create datasets that aggregate services to the administrative level, and this information can be helpful in reporting and measuring financial-inclusion metrics such as the number of services per 10,000 head of population. This allows a limited set of use cases around measurement to be met relatively cheaply and easily, assuming that the address data is available and accurate.

## Challenges

Until all financial service providers are able to publish high-quality, regularly updated address data online, this method will produce incomplete and unreliable data for decision-making for the majority of FSPs. One of two things would have to happen to make this form of data collection truly useful:

- A complete, accurate and up-to-date address database linked at a spatial dataset, also known as a cadastre, cadastral map or cadastral survey, together with the publication of the addresses of all financial services, or
- The publication of complete, accurate, up-to-date coordinate data by financial service providers on their website

In the first case, the investment that is required to maintain an up-to-date cadastral map is huge. The UK postal service spends tens of millions of pounds every year to keep its database up to date, and is simply beyond the means of many countries.

In the second case, if financial service providers were to get to this level of data collection and publication, without intervention or guidance from market facilitators or regulators, then this method would be an extremely efficient and cost-effective way to consolidate the data. This, however, seems highly unlikely.

### Most useful for:

Producing a relatively quick and cheap aggregated overview of services. For example, addresses of services taken from financial service provider websites could be used to create calculate basic financial inclusion metrics, such as the number of service points per 10,000 population.

However, given the serious limitations of the data produced by this methodology, it is not one that is currently recommended for developing markets. It could be used effectively in developed markets, where all providers publish the locations of their services and where the country has a standardised street addressing system that allows accurate address geo-coding to the building level.

## Handheld GPS devices and pen and paper

Although this method is used far less frequently than the others mentioned here, and with good reason, it is worth mentioning in order to highlight why it is not a suitable methodology for this kind of data collection. Versions of this method have been, and are, being used by both traditional and digital financial service providers to capture the locations of their services; and in nearly all cases, the data produced has been unusable.

### Methodology

A dedicated, handheld GPS unit is used to capture the locations of financial access points in the field. The points are sometimes saved on the device itself, but more often the coordinates are copied from the device screen onto a paper data-collection form, which is also used to record any required supplementary data.

### Benefits

Dedicated, handheld GPS units (such as the one pictured here) provide very accurate GPS positions, which – in the more expensive units – can be accurate to within a few centimetres. This makes them suitable for land survey work and other uses that require a very high degree of accuracy.

## Challenges

Handheld GPS units are often not very user-friendly and can require a significant amount of training to be able to use properly. They also have an array of complex and specialised features, which are not required for this kind of data collection exercise. This can lead to a situation where an inexperienced user can accidently change a vital setting without knowing it. This can, in most cases, make the dataset unusable.

Using a handheld GPS and paper data collection form requires data collectors to accurately copy down long strings of digits and decimal points from a usually tiny screen, often in bright sunlight. This results in a very high percentage of transcription errors. Depending on which digit is copied incorrectly, this can lead to an error in the data of a few metres or an error of a few thousand kilometres and once the error has been made it is almost impossible to correct. Not only does this method create the opportunity for transcription errors in the field but, once the data has been collected on paper forms, it needs to be transcribed a second time into a spreadsheet, database or another digital format. This is not only prone to the same kinds of errors but is also costly and time-consuming.

## Summary

This is the only method, listed here, that – without equivocation – should not be used to capture the locations of financial services outlets.

“Using a handheld GPS and paper data collection form requires data collectors to accurately copy down long strings of digits and decimal points from a usually tiny screen, often in bright sunlight. This results in a very high percentage of transcription errors.”

# How to find us:

## Get involved. Contact us.


**David Taylor**
**E:** david@usabledata.co

**Grant Robertson**
**E:** grant@i2ifacility.org

**Damola Olowade**
**E:** damola@i2ifacility.org

**Petronella Tizora**
**E:** petronella@i2ifacility.org

+27 11 315 9197
**i2ifacility.org**

**@i2ifacility**
**/insight2impact**
**/insight2impact**
**/i2ifacility**



Established by

cenfri

FMT FINMARK TRUST

Sponsored by

BILL & MELINDA GATES foundation

mastercard foundation