

MIX Microfinance World: Mapping Africa Financial Inclusion- Data Methodology

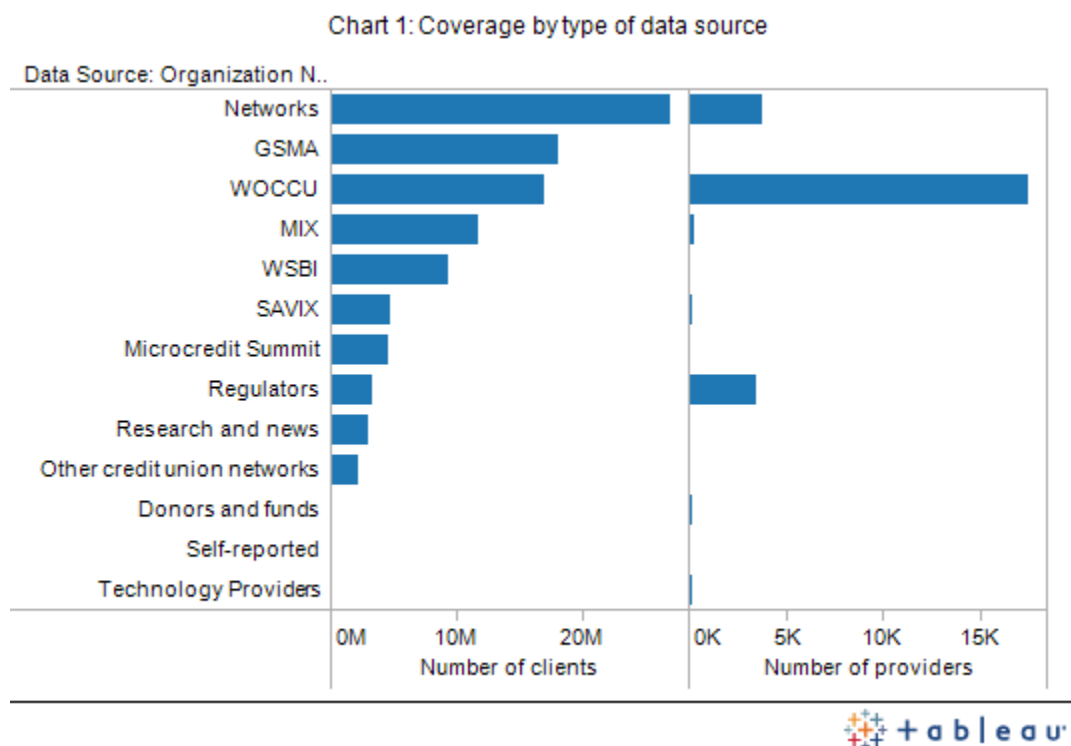


Methodology

In this section, we describe the methodology of our research in more detail, along with some conclusions based on the process behind constructing this data set.

We began by compiling data from several different sources, including local and international networks, regulators, research organizations, donors, investors, news sources, software providers and others. Overall, we referenced 64 distinct aggregate data sources, as well as many individual providers, spanning over 5600 distinct observations. We were often able to rely on publicly-available data in most markets, which speaks to the increasing support for transparency from sub-Saharan Africa.

Local knowledge is strongest: A diverse set of data sources allows us to take a broad view on each market and to cross-check data from different providers. The types of sources, with total clients and institutions is below in Chart 1. (These totals we calculated before removing duplicates and double-counting, which best allows us to see the depth as sources of raw data.) Local networks cover the largest share of clients, which supports recommendations to localize supply-side estimates: local actors have strong interest in providing timely and comprehensive data.¹ We can build better data resources and minimize duplication by leveraging this work.

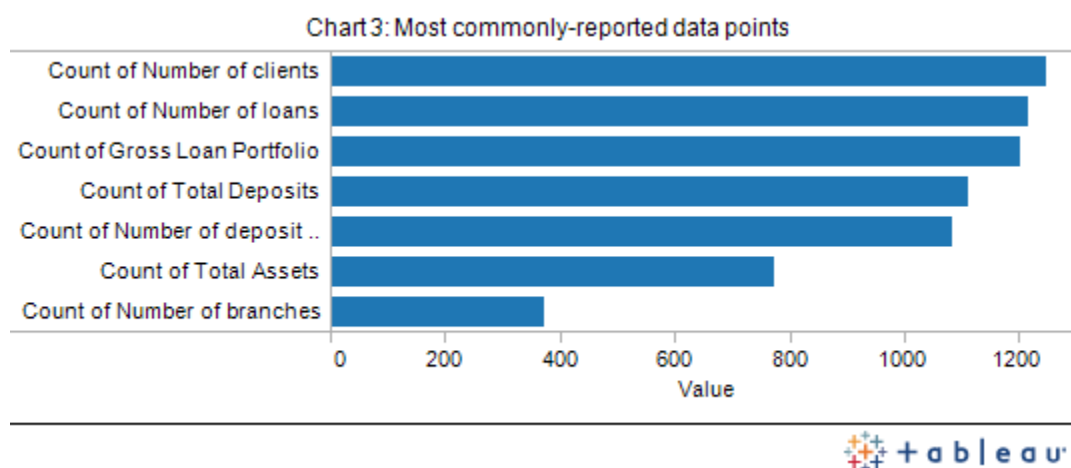


[ALAFIA](#) in Benin provides a strong example of how deep local data sources can be. This network tracks 37 institutions in the sector from 1998 to the present. While the MIX Market site tracks the largest providers, if we compare aggregate figures from MIX Market to data from ALAFIA, we can see that both

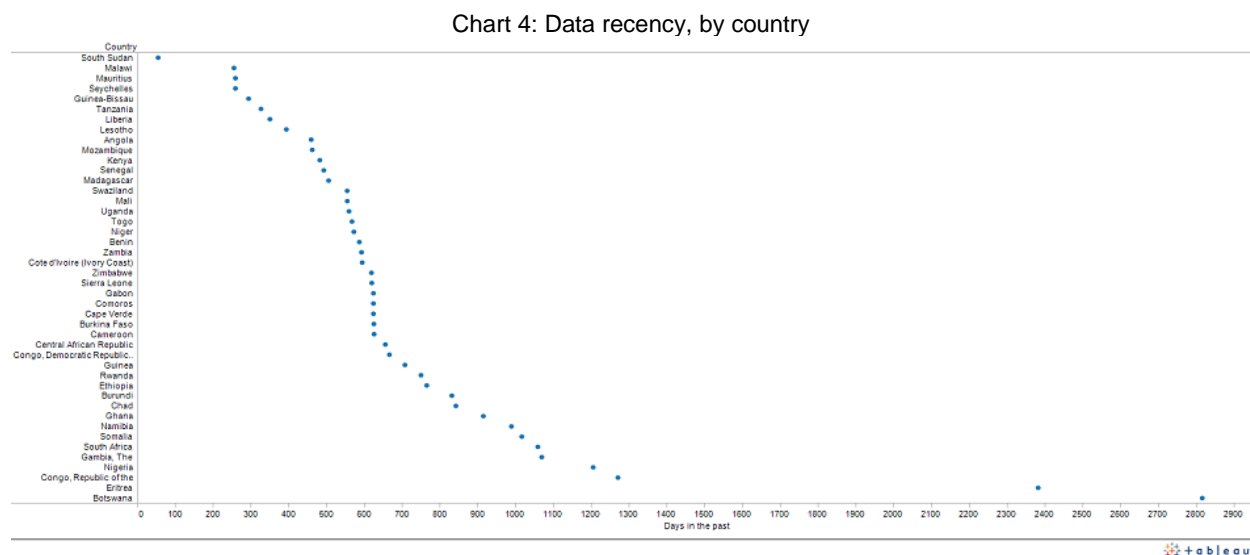
the levels and the trends differ over time. Deeper data is [still accessible via MIX Market](#) for investors or others, but the landscape figures provide better market context.

In many cases, different sources provide data for the same providers. In some cases, different data sources provide data at different points in time. In these cases, we have taken the most recent estimate. In other cases, different sources cover the same provider for the same point in time. In these cases, we have relied on country-level coverage to choose data sources. A subtler type of overlap occurs when we have aggregated and disaggregated data with overlapping coverage. In these cases, we have used information on the linkages between different actors to remove double-counted entries and to allow simple aggregation.

We tracked only a small number of indicators for this effort, following the data provided directly in the source. Chart 3 showing the distribution of the raw data points is below. This confirms the fact that monetary indicators and credit indicators are reported more often (or by more sources) than client indicators and savings indicators.² Branch locations are among the least-often reported, making detailed geographic mapping a challenge in many markets, although there are instances that permit this [level of detail](#).



How up-to-date is available knowledge on Africa? With data from the various sources in a single platform, we determined the most recent data point for each provider. Often, data was available within the last year, but in some instances, data was published with less frequency or was less readily accessible. Chart 4 shows the weighted average ‘recency’ for each market (computed as a weighted average by the number of clients and the number of days in the past). Across all markets, the median number of days (from the date of this report) is 620 days, which puts the ‘typical’ data point in late 2009. Fittingly, [Africa's newest country](#) also has the most recent data.



Markets with the least recent data likely deserve the most attention in terms of data collection and distribution. Markets with less-frequent reporting tend to be either post-conflict markets or those that lack strong local networks or regulators for data collection. Of course, if more recent data is available for any provider we can incorporate it and update the data and mapping. (Conversely, when we have information that a provider or group of providers is no longer in operation, we remove their data from the list.)

What information is missing? Not all sources provide the same data points. Regulators may focus on monetary figures, rather than client or account data. Resources focused on credit may provide more credit metrics; resources focused on savings better track deposit metrics. We thus need a method to estimate missing data in order to aggregate and compare data across these varied sources: if a data source reports partial information, can we infer the missing points? Fortunately, we can use benchmarks from MIX Market and other contributing data sources to develop realistic estimates for these missing data points. While other efforts have focused on econometric models calibrated across multiple markets, we use local estimates grouped by country, legal status and time whenever possible in order to target specific, relevant comparable figures. Overall, we have employed six rules for estimation of missing values:

Rule	Description	Notes ³
a	If number of clients is missing, and either number of borrowers or number of depositors is present, estimate number of clients as max(borrowers, depositors)	This rule is likely to under-estimate total clients as it assumes that savings and credit client groups completely contain one another.
b	If number of clients is numeric value, and number of savers is missing, and provider type is credit union or savings group, use number of clients = number of depositors	We assume that all members of savings groups and credit unions must open a savings account; if this is not true, we will over-estimate deposit outreach at these providers. We exclude banks and other providers that may mobilize savings from this group as we cannot infer that all clients open deposit accounts. Thus this rule is likely to under-estimate savings outreach from these other providers.
c	If number of depositors is numeric value, and number of borrowers is blank and charter is credit union / cooperative, NBF / NGO or bank, use depositors / borrowers benchmark ratios to populate number of borrowers	For this rule, we match the benchmark ratios of depositors to borrowers by country and charter status and year. Reference data is available here (for MIX charter types); for savings groups, ratios are based on the SAVIX site . It is not clear if this over or under-estimates credit outreach.
d	If number of borrowers is numeric value, and gross loan portfolio is not, and provider type is among MIX categories, then use average loan balance / GNI per capita for country in question for year in question	For this rule, we match average loan balances based on MIX benchmarks, for which a sample report is here . Given that the MIX data comprises mostly microfinance-focused institutions, it is likely that these figures under-estimate total credit volumes, since loan balances are smaller for these institutions.
e	If number of depositors is numeric value, and deposits volume is not, and provider type is among MIX categories, then use average deposits balance / GNI per capita for country in question for year in question	Matching criteria and reference benchmarks are the same as described above, as is likely bias.
f	If loan or deposits volume is numeric value, and loans or depositors is not, use reference benchmark for savings banks to estimate number of loans / deposits. Re-apply rule a) after this adjustment.	We constructed benchmarks for savings banks for 2009 based on WSBI members' data, as linked here . As above, average balances / GNI per capita were used and converted to dollar amounts using local GNI per capita levels.

Most of the rules are either likely to underestimate outreach and volumes or do not have clear bias. The estimation methodology should thus be sufficiently conservative to deploy for this effort without inflating country and region totals. In addition, this straightforward methodology allows other practitioners to replicate and extend the results.

Pulling the data together Once we have gathered data from the various sources, found the most recent data points, estimated missing data and removed duplicates and overlaps, we have a final scrubbed data set ready for presentation. A link to this raw data set is posted [here](#). This level of detail allows for top-to-bottom review, and lets users dig into the details, beyond what is available in a single report. We provide this to allow policy-makers and practitioners to disaggregate high-level figures to individual institutions and financial services providers. In addition, this supports validation and verification of the data set for future improvements.

Related Articles:

- [Overview](#)
- [Historical Efforts and MIX's New Approach](#)
- [Results Review and Next Steps](#)

1: In several cases, local networks and regulators collaborated to provide or publicize data, and their contributions may be best considered jointly.

2: Data on PAR > 30 was captured in some markets, but omitted from this chart and the final results due to limited availability.

3: For all rules, if a match for country and legal status and year is not available, we will use an adjacent year if available. Failing that, we will use an all-Africa benchmark for the same year. All average balance metrics are based on figures relative to GNI per capita first, and are then converted to dollars based on local GNI per capita levels for the same year.