insight2impact

Advancing Financial Inclusion

Capturing and recording spatial data

Guidelines, standards and best practices

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

Capturing, managing and analysing spatial data used to be the preserve of specialists who had training that allowed them to effectively use complex hardware and software technologies. With the recent rise of the smartphone and location intelligence platforms that allow users to analyse, style and share spatial data more easily, spatial data is becoming more powerful and is open to a wider audience than ever before.

While the dependence on specialised skills is becoming less important and the spatial world is opening up to new users, it is important that non-specialist users understand enough about the capture and recording of spatial data to ensure high-quality results. This note lays out some of the issues to be aware of when capturing and recording spatial data, offers guidelines and suggests standards and best practices.

As with any data-driven product, precision and standardisation are key to achieving a high-quality result. Details such as where people stand when they are capturing Global Positioning System (GPS) coordinates, how the coordinates are formatted and how they are communicated can seem trivial, but they can have a dramatic effect on the quality of the data produced.

Recording spatial data

Coordinate formatting

Coordinates can be formatted in a range of ways. The table below shows a variety of ways in which the same coordinates can be presented.

Table 1: Presentation of GPS coordinates

Format	Latitude	Longitude
Degrees, minutes, seconds	6°48'0.4932"S	39°18'52.4736"E
Degrees, decimal seconds	6°48.008220'S	39°18.874560'E
Decimal degrees	-6.800137	39.314576

Northing and easting
Usually measured in metres, together with the Universal Transverse Mercator coordinate system

Format	Northing (y-coordinate)	Easting (x-coordinate)
Degrees, minutes, seconds, UTM zone 11	50°52'10"N	115°39'03"W
Shortened northing, easting pair	11U 594934	5636174

While it is not necessary to go into detail about the differences between these systems, it is important to be aware of their existence and to be able to spot them. Using data collection apps or bots on smartphones will, to a large degree, remove any problems that may be caused by differing coordinate formats. This is because they tend to use standard latitude and longitude formatted as decimal degrees, with the coordinates being labelled "latitude" and "longitude" and being stored in separate columns of a spreadsheet or database. It is recommended that all spatial data be collected and stored in the format below.

Latitude	Longitude
-6.800137	39.314576

If reporting systems are being used (which allow service providers to import their data into a common or shared database), the data should be checked to ensure that it meets the following standards:

- It is formatted as a decimal, where there are no more than two digits before the decimal point.
- Both latitude and longitude values can be either negative or positive.
- It is possible to draw a box around the country being mapped, which allows the setting of upper and lower limits on the latitude and longitude values that can successfully be imported.

Coordinate systems

Coordinates should be captured and recorded by means of the coordinate system *World Geodetic System 84*, commonly known as WGS 84 and having EPSG:4326 as reference for GIS software. WGS 84 is a standard reference coordinate system that is used by the Global Positioning System and has become the standard for smartphone-based and web-mapping location services.

Spatial data captured using standard open-source software, such as Open Data Kit (ODK)[1], will automatically be captured using WGS 84.

GPS accuracy

When using app-based data collection tools for smartphones, it is possible to set an automatic minimum accuracy level for capturing spatial data. This feature allows GPS coordinates to be logged automatically when the accuracy level falls below a predefined level, measured in meters.

The trade-off, when deciding what level of accuracy to set by default, is between time and accuracy. One of the biggest factors in obtaining an accurate GPS reading is the time that is allowed for the phone to obtain a fix on the available satellites. The more satellites there are available and the less obstructed the view of the sky, the quicker the fix is likely to be.

In most cases, a default accuracy of 10 metres provides high-enough quality of data to identify the specific service point while being a reasonable requirement for most areas that are likely to contain financial access points. A 10-metre accuracy level is more than enough to conduct relevant spatial analysis for policy and business use cases and is also high enough to allow the data to be used to inform customers of the location of the service point.

Data collection tools should allow users to manually override the requirement for accuracy below 10 metres in cases where atmospheric conditions or other factors make it impossible to capture more accurate data.

Data on the accuracy of the GPS coordinates of all points should be captured in the metadata and stored in the data management system.

“One of the biggest factors in obtaining an accurate GPS reading is the time that is allowed for the phone to obtain a fix on the available satellites. The more satellites there are available and the less obstructed the view of the sky, the quicker the fix is likely to be.”

1 https://opendatakit.org

Where to stand when capturing GPS coordinates

It is important to stand in an area that has a clear view of the sky in order to be able to capture accurate GPS coordinates.

GPS coordinates for structures that provide financial services should be captured a metre or two directly outside the main customer entrance.

An exception to this would arise where a bank or other service point (such as an ATM) is located within a shopping centre or other building complex, which means that the main entrance does not connect directly with the outside world. In this situation, the GPS coordinate should be captured at the main customer entrance to the building complex, and a note should be made in the address or notes fields.

Some data collection applications (such as Open Map Kit) use the GPS device to narrow down the area and then allow the user that is capturing the location to define the exact location by placing a pin on top of high-resolution satellite imagery of the area. While this can produce high-quality results, the quality of the data will be highly dependent on the knowledge, skill and training of the individual that captured the location. While both techniques are valid and have their own advantages, it is important to choose the correct tool for the overall data collection approach.

One location, multiple services

The rise of digital financial services has resulted in interoperability where a single service point (a merchant or an agent) can offer products from multiple financial service providers. This has led to a situation where it is important to standardise how the location information for these points is captured and managed.

Ideally, only one set of GPS coordinates should be associated with any given service point.

The challenge in standardising the spatial information for points that provide multiple services arises when service providers are asked to self-report the locations of these services. If a kiosk provides mobile services for Telco 1 and Telco 2 and both companies are asked to collect and report the spatial data for this service point, then the GPS points that are reported will overlap but, given a likely margin of error of around 10 metres in any direction, the GPS points will not match exactly. This means we cannot use the GPS coordinates to say with any degree of certainty whether the location captured by Telco 1 and the location captured by Telco 2 are the same location. We could use the business name, but (given the challenges that will arise with enumerators capturing names with slight differences in spelling and businesses that are in close proximity with similar names) this is not a practical solution – especially in markets with hundreds of thousands of agent locations.

Without a standard way to link data from different providers to a single service location, we risk either dramatically under- or over-counting the number of service locations in existence. For example, if the service point above were mapped by both Telco 1 and Telco 2 and both use slightly different spellings of the name, this single service point could easily be counted as two service points by a regulator when aggregating data from different providers. It is easy to see how, in markets where one service point could provide services for three or four service providers, the number of service points could easily be dramatically over-estimated.

In some markets (such as Nigeria), a unique ID is assigned to agents when the first application is made for that location to provide financial services. Any other service providers who subsequently wish to add services to the existing location then report the addition of new services by means of this unique ID. This has the added benefit that the spatial data for any location only needs to be captured once, when the first financial service starts operating in a location.

If no suitable coding system exists that can be exploited, the regulator should create a standardised coding system to provide service point locations with a human-readable, unique identifier, i.e. a code “12345678”, which is displayed at the service point. Service providers and regulators can use this code to accurately aggregate different services and tills to a single location. Customers can use it to identify outlets, and as a reference when using complaints recourse mechanisms available with regards to reporting fraud or other abuses to regulators and providers. The same code could also be made available in a machine-readable format (i.e. a barcode or QR code) that can be scanned using a smartphone or another data collection tool.

Transcription

Transcription is the process of manually copying data from one source to another. In this instance, the most common times that transcription is likely to occur is in the field, when individuals may be required to copy GPS coordinates from a smartphone or GPS device onto a paper form. If this happens, a second level of transcription will be required to enter the handwritten coordinates into a spreadsheet or other digital medium. This could be done either by a human or by Optical Character Recognition (OCR), which uses a scan of the text and a program to identify patterns to turn the hand-written text into digital text.

Because of the high potential for error, any use of transcription is strongly discouraged. The error rates of human transcription and most OCR software are too high to be acceptable. GPS coordinates are particularly susceptible to errors in this respect, as there is no context to calibrate errors. For example, mistyping one letter in a word would probably allow the reader to infer the correct word from the context. However, GPS coordinates have no context, and an error in a single digit (e.g. the simple confusion of a one with a seven) could result in an error of several thousand kilometres in the final data.

Methods for capturing data

Geocoding

Geocoding is the process of turning addresses into GPS coordinates. Effective geocoding requires two things: high-quality address information for existing financial services and an accessible, accurate, up-to-date spatial database for street addresses, also known as a cadastre. The advantage of this approach is that, in theory, it would allow the creation of high-quality spatial datasets from the street addresses of the service locations.

In reality, the challenge is that the creation and maintenance of an accurate, up-to-date cadastre is both expensive and time-consuming. Very few countries have the resources available to do this effectively; and, as a result, it is usually only possible to geocode addresses in capital cities and major urban centres. Even in highly urban areas, a lack of standardised spelling of road names, as well as the existence of roads and settlements that may not be officially recognised, or that simply do not have names, mean that the quality of the geocoded data is often questionable, at best.

This methodology is only really suitable for creating highly aggregated data based on broad administrative areas.

Remote sensing

Remote sensing is the technical term for identifying something of interest using aerial or satellite imagery. Many financial service providers have used this technique to identify their branch or other service locations using imagery on services like Google Maps or Google Earth. Once locations have been identified, these services allow them to be saved and the GPS coordinates can then be extracted.

While this technique may be viable for a commercial bank with a relatively small number of easily identifiable branch locations in urban areas which have good satellite imagery and street data, it can be a real challenge to identify services such as mobile money and bank agents who may operate out of much smaller premises that may not be easily identifiable using satellite imagery. The quality of free and publicly available satellite imagery also varies dramatically between highly populated urban areas and sparsely populated rural towns. These challenges seriously limit the potential of remote sensing to provide a comprehensive solution to capturing the locations of financial services.

“The challenge is that the creation and maintenance of an accurate, up-to-date cadastre is both expensive and time-consuming. Very few countries have the resources available to do this effectively; and, as a result, it is usually only possible to geocode addresses in capital cities and major urban centres.”

Apps and Bots

Currently, the most reliable way to capture high-quality spatial data for financial access points is using a smartphone. Smartphones are set up to work using the spatial standards and formats listed in Table 1. They also allow for checks and validation of supplementary data to be performed in the field. For example, a smartphone data collection app can check that a phone number does not contain letters and is the right length, it can also standardise the names of administrative units using dropdown menus instead of allowing free text to be entered.

Apps

There is a growing number of smartphone-based data collection apps and many of them take advantage of free and open source libraries, such as Open Data Kit[2]. These apps allow surveys to be designed to capture a range of data types including location data, free text, single and multi-select questions, images and scans of bar or QR codes.

Bots

Bots use existing messaging services, such as Facebook Messenger, WhatsApp, Viber and WeChat, to construct standardised conversational surveys, which users interact with as if they were talking to a real person. As users are not required to download a new app or become familiar with a new user interface they can significantly reduce the need for training to be able to collect the data. The ease of use also makes this a particularly interesting option to allow interacting with the public, for example, as a channel for users to report incidence of fraud to a regulator.

Both apps and bots allow for the collection of high-quality spatial and supplementary data through the use of questions with standardised responses that require minimal interpretation from the user.

“Both apps and bots allow for the collection of high-quality spatial and supplementary data through the use of questions with standardised responses that require minimal interpretation from the user.”

2 https://opendatakit.org

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