

---

# **The reference framework for first mile farm data (Beta)**

*Release beta*

**The reference framework for first mile farm data (Beta)**

**Dec 25, 2017**



---

## Contents

---

<b>1</b>	<b>1 about the reference framework</b>	<b>3</b>
<b>2</b>	<b>2 Use cases</b>	<b>5</b>
2.1	2.1 Use cases for farmer groups . . . . .	5
2.2	2.2 Use cases for voluntary standards systems . . . . .	5
2.3	2.3 Use cases for certification bodies . . . . .	6
2.4	2.4 Use cases for technology providers . . . . .	6
2.5	2.5 Use cases for monitoring and evaluation . . . . .	6
<b>3</b>	<b>3 The data model</b>	<b>7</b>
3.1	3.1 Approach . . . . .	7
3.2	3.2 Overview of model JSON model . . . . .	8
3.3	3.3 Schema details in JSON . . . . .	9
<b>4</b>	<b>4 Extensions</b>	<b>13</b>
4.1	4.1 MARS adoption observations . . . . .	13
4.2	4.2 CocoaAction . . . . .	13
<b>5</b>	<b>5 Uniquely identifying data elements</b>	<b>15</b>
5.1	Key message . . . . .	15
5.2	5.1 Introduction . . . . .	15
5.3	5.2 The InternalID . . . . .	16
5.4	5.3 The GeoID . . . . .	16
5.5	5.4 GlobalID . . . . .	17
5.6	5.5 Alternative ways of identification . . . . .	17
<b>6</b>	<b>6 Spatial data acquisition and standards</b>	<b>19</b>
6.1	Key messages . . . . .	19
6.2	6.1 Introduction . . . . .	20
6.3	6.2 Using a Navigational Satellite System to collect spatial data in the field. . . . .	20
6.4	6.3 Using a Geographical Information System (GIS) to collect spatial data. . . . .	22
6.5	6.4 Some more background on the complexity behind geodata . . . . .	24



```
.wy-table-responsive table td { /* !important prevents the common CSS stylesheets from overriding this as on RTD  
they are loaded after this stylesheet */ white-space: normal !important; }
```

```
.wy-table-responsive { overflow: visible !important; }
```

→



---

## 1 about the reference framework

---

This is the development version for the reference framework for first mile farm data. This resource aims to evolve into a set of shared principles on how to collect, store and exchange data about farmers, farmer organisations, farms, plots and other data related to agricultural practice.

The term ‘the first mile’ refers to the position of farmers at the start of the agricultural value chain. When moving up the chain, following all steps towards consumption the distance to these food production area’s literally increases. While transparency and traceability in the value chain is of growing importance, data collection about ‘the first mile’ is generally difficult due to logistical reasons. There are many food producers and they are generally located in more remote areas making data collection labour intensive and expensive, especially in the case of small holder farming.

The purpose of the reference framework for first mile farm data is to align the way in which data about farmers, farms and farm practices are collected and stored by different organisations to facilitate the data exchange and interoperability of first mile farm data sets.

The current version provides elements:

1. **A generic data model for first mile farm data.** by conceptually modelling the relations between the core data entities such as fields, farms, farmers and farmer organisations for first mile farm data collection. This generic data model can be used as a template for organisations to structure first-mile farm data collected in the field. The model can also function as a neutral data model for data exchange between organisations. The generic data model is designed in such way that it should be relatively straight forward for other organisations to make a mapping from their own internal data structures to generic first mile data model. The data model is available in JSON format and in xls format in GitHub.
2. **Extensions.** The first mile farm data model can be easily be expanded with new data attributes or even new data entities for different purposes. This is demonstrated in the section with the extensions where examples are added for Cocoa action program and MARS Adoption Observations. For the 2 programs the data formats are available as together with protocols on how to collect the data in the field. By adding more and more extensions to the framework, a central repository emerges of proven data formats and their data collection protocols. Organisations can benefit from this repository harvesting the data formats and data collection protocols they need, contributing the standardization and interoperability of first-mile farm data collection in the field.
3. In addition a approach **to uniquely identify first mile farm data globally** using a system of 3 identifiers is presented and best practice recommendations are given for **spatial data collection** in the field.

These three elements provide the basis for the reference framework. Other organisations are invited to help extending the framework by adding the data standards and data collection protocols. To add or to ask questions about the first mile reference **please contact: [andre.jellema@data-impact.com](mailto:andre.jellema@data-impact.com)**.

The development of this framework has been developed by [UTZ](#) supported by [Data-Impact.com](#) and [Open Data Services](#) and is funded by [ISEAL](#)

The following organisations are thanked for their reflection and contributions to the framework:

- Rainforest Alliance
- Grameen foundation
- AKVO
- SupplyShift
- SourceTrace



This chapter provides a nine of use cases describing where the reference framework and its data model applies.

#### 2.1 Use cases for farmer groups

*Narrative 1: to support the design of member management systems.* Smallholder farmers in many sectors and countries are organized into cooperatives or farmer groups. Farmer groups often use management information system to store and retrieve information about their members. The collection and storage of this information happens more and more by digital means, also when smallholder farmers are involved. Farmer organisations use this information for decision making and to provide services to the farmers. In the management information system many different data elements may be stored including: personal information, demographics of the farmer group, expected and previous harvest results, the use or need for inputs, information on pest and diseases, soil quality etc. The data model as provided in this document provides a means to: 1 structure data in such a management information system, 2 exchange data between systems and 3 provide recommendations on how data can be best collected in the field.

*Narrative 2: to facilitate data exchange between member management systems.* Farmer groups are dynamic. Farmers often have different options to join a farmer group in their region. When a farmer switches group, it would be beneficial if the 2 farmer groups involved can easily exchange their data. This includes personal information, as well as the location of the farm description and the historical farm performance. The reference framework provides a format that facilitates the exchange of data between organisations if both organisations comply to the format or map their data infrastructure upon the proposed data format.

#### 2.2 Use cases for voluntary standards systems

*Narrative 3: to facilitate farm data exchange.* Voluntary standards systems generally work with many different farmer groups in different commodities around the world. As a result of the digital revolution in agriculture, more and more first mile data about the farms is being shared between the farmer groups and the voluntary standards system. The complexity of re-using farmer field data within the standards organisation will reduce if the farmer groups are aligned in the way they digitally collect and store their farmer information using this framework.

*Narrative 4: to support development of innovative ways of compliance evaluation.* Voluntary standards systems are experimenting with new ways to evaluate the compliance to their standards. Examples include the application of earth observation data and (geospatial) data analysis. Accurate and standardized spatial data on field locations as well as ground observations are required to do this analysis efficiently and to develop automated methodologies.

*Narrative 5: to support the multiple certification.* If the required data collection by farmer groups, companies and auditors between different standard organisations would be aligned, switching between standards or adopting multiple standards would become easier and cheaper.

*Narrative 6: to help avoiding double selling of multiple certified produce.* Double selling is the practice of a farmer to sell the same quantity of produce twice with different certification labels. This may happen if an amount of produce is certified for 2 or more standards. On the first sale the farmer may sell all farm produced certified product as label A. As a result the farmer can no longer sell produce labelled A, because of registration at voluntary standards system A. However voluntary standards system B may not be aware of the sale of the produce. In principle, the farmer could buy the same amount of uncertified produce on the market and sell this as B certified produce. A data standard facilitating easy exchange of farm data between different voluntary standards systems helps to finish this kind of practices.

## 2.3 Use cases for certification bodies

*Narrative 7: to ease audits.* The role of a control body is to inspect if all practices at a farm or farmers group are being performed conform certification requirements and to inform the voluntary standards system about the results of the inspection. If the auditor, the person who actually performs the inspection, could receive farm data in advance, he could make a pre analysis which farms he likes to visit and can use this information as a guidance in the field to focus on specific farmers or topics. The data can also be helpful to validate some audit points in advance before the field visit or to pre fill some of the data points that need to be collected in the field, for example it is less work to check a field boundary then to measure it, saving time and money. The reuses of existing data will be facilitated if all organisation use the same data formats and use the same data collection methods, or if they could map their data on a standard data exchange format.

## 2.4 Use cases for technology providers

*Narrative 8: to align the technology demand from clients.* If the clients are more aligned in what data need to be collected, how data is collected and how data is stored, it will become more easy to develop the specific tools. This applies to data collection tools, management information systems, data analytics, visualization and data exchange via an API.

## 2.5 Use cases for monitoring and evaluation

*Narrative 9: re-use of data and meta analysis.* Numerous projects are run by NGO's, companies, voluntary standards systems to improve the life of small holder farmers. It is often challenging to assess the impact on the different farmer communities. This reference framework will ease the exchange of farm data between organisations and project. Researchers could reuse the data making a more thorough analysis of impact and to compare between projects, regions or even in time, leading to more effective interventions.

---

### 3 The data model

---

#### 3.1 Approach

In the [three schema approach](#) a conceptual data model is proposed as an effective way to integrate between different databases and organisations. In the conceptual model the data entities are defined in the way that the end-users think and talk about concepts in the real world. It is therefore assumed that multiple organisations can easily develop a mapping from their own internal data structures onto such conceptual model because of the internal logic of the end-user community. The conceptual model can therefore function as a neutral and organisation independent interface for the data from 1 database to another database. If all organisations would develop such a mapping (or adapt their own internal data structures), effortless and automated data exchange and data integration can take place. The reference framework for first mile farm data aims to put the foundation for such a neutral model.

In ‘first mile’ projects many different aspects of farms are being collected. Examples are the agricultural and economic performance of a farm, farming activities performed at a farm, social and environmental conditions, compliance to standard measures etc. It is beyond the scope of this document to address all of these aspects in detail, including the recommended field data collection methodologies and associated data formats. This document presents a generic structure by representing the core components of a farm as data entities. It will only discuss data attributes that are essential for the internal structure and which are needed to be able understand and combine the data, such as farmers ID or the shape of the fields or are commonly collected such as the address or expected yield. It will also present the logic how the model can be expanded with additional data attributes or data entities providing the option to customise the model and to provide the opportunity to expand the generic model into a more all comprehensive model at a later stage.

##### 3.1.1 The conceptual model

To define the conceptual data model, the concept of a farm is taken as the starting point. A farm in this reference framework is defined as a collection assets managed as 1 entity with the primarily objective to raise living organisms for food or raw materials. At the farm crops are cultivated on one or more plots of land and farm animals are being kept. The farm is run by a farmer, often supported by farm workers. Different buildings may be present at the farm to support the work. The farm may consist of multiple plots, herds or buildings. Workers are not bound to 1 farm but may work at different farms. A farmer may manage several farms. A farmer may be member of one or more

farmers groups. A farmer group can be a cooperative or a group of farmers loosely organised around a government or corporate program.

Many organisations are making observations in a plot during their field work. The data entity 'observation' is introduced to map this observation data.

The conceptual model is flexible to include a number of different farmer and farm types depending on the ownership of the farm and ownership of farm assets. Traditionally one may think of the farmer being the owner of both the farm and the farm assets, however this is not always the case. In many cases a farmer will hire or lease part of the production assets. 'Sharecroppers' or 'Tenant farmers' make use of agricultural assets which they do not own and have to pay a percentage of the profit, part of the produce or both. In 'systems with communal land rights' members of the community are getting fields assigned to use for farming, but do not own the land. Land use rights may change through time. The land of a farm is therefore a not static asset but may change through time.

In some cases the farmer is more like 'the manager' and not 'the owner' of the farm benefitting directly from the produce of the farm. When the rights are shared with a group of individuals, the farm can be called a 'Communal farm'. When ownership rights are with a company the farm can be called a 'Corporate farm'. By managing the concept of ownership well for the different data entities all of these different type of farms and farmers can be included in the model, even while ownership changes over time.

The logic of the model is to map the first mile data elements collected in the field work at the right level data entity. Eg.:

- The certified stock of a cooperative should be a data element of the data entity group
- The total certified produce of a farm should be a data element of the data entity farm
- The expected harvest of a field should be a data element of the data entity group

The dark blue elements in the figure above indicated in the model will be detailed out more below. The lighter blue elements to visualize the logic in the model and to indicate possible extensions in the future or customization options.

## 3.2 Overview of model JSON model

The complete data structure is visualised in the table below:

- All required data attributes are indicated in bold,
- By clicking on the blue table title -> all data attributes become visible,
- By clicking on the data entity buttons -> all data attributes become visible of that data entity and,
- By clicking on the { } symbols -> the JSON becomes visible

In [JSON schema language](#) the data entities making up the conceptual model (Group, Farmer, Farm, Plot, Observation) as described in the conceptual model are stored as [arrays of JSON objects](#).

Each time a data entity is recorded a new JSON object is created and added to the array. The recording of the same data entity at another point in time needs to be added as a new record to allow for continuous monitoring and to allow the composition of farms and ownership to change overtime. The recording of the data entities therefore are timestamped, this takes place in the GlobalRecordID object (see below) In the reference framework a system of 3 identifiers is proposed to uniquely identify objects and model the relationships between these objects. These Identifiers are called: 1 the GlobalRecordID, 2 the GeoID and 3 the InternalID. The system allows the exchange and merging of data from different organisations if both datasets are following the first mile data structure.

- **The GlobalRecordID** is a globally unique identifier for First Mile Farm data record.
- **The InternalID** provides a means to internally structure the data set, uniquely identifying the objects in the database and the relations between them.

- **The GeoID** provides an imperfect means to uniquely identify the object in the field and to clean the dataset when it is combined with datasets from other sources.

This mechanism will be explained in more detail in section on ‘[Uniquely Identifying Data Elements](#)’.

It is not required to use all data entities (Farmer group, Farmer, Farm, Plot, Plot Observation) in a databases while using the data structure. Some organisations only collect plot level data, so no details are needed/available on which farmer manages the plot or to whom the plot belongs. The arrays for group, farmer, and farm can be left empty.

However because all data elements are mapped on the same data entities, datasets with different levels of ‘completeness’ can still be combined for analysis. E.g. if an organisation who collects plot data, farm data and farmer data uses the same structure to organise its plot data as an organisation who only collects plot data, all plot level data entries can be combined and analysed collectively in a meta analysis.

As a result, to maintain compatibility for different global data sources, with different combinations of data entities (Farmer group, Farmer, Farm, Plot, Plot Observation), it is essential that all data attributes are mapped at the right data entity level. For example it does not fit the logic of the reference framework to add an attribute ‘farm size’ to the data entity ‘farmer’, even if that farmer owns or manages that farm. The farm and it’s size are inextricable, even if this is the only attribute recorded about farms in the data set. As a result the farm size of a farmer’s farm can only be found by first finding the farm instance of the farmer’s farm using internal keys and next the size of that farm instance.

### 3.2.1 Extensions

The reference framework is designed in such a way that new data attributes or even new data entities for different purposes can be added to the model easily. This is illustrated under the heading extensions, for Cocoa action program and MARS Adoption Observations. The data formats are available together with the official protocols on how to collect the data elements in the field.

By adding more and more extensions to the reference framework, a repository emerges of ‘proven’ data formats and data collection protocols. Organisations can benefit from this repository harvesting the formats and data collection protocols they need for their own data management. Having different data formats and data collection methodologies in one repository will also facilitate further standardization and interoperability discussions. To add to the reference framework please contact [andre.jellema@data-impact.com](mailto:andre.jellema@data-impact.com).

In the following sections the core data entities of the generic data structure for farm data collection, storage and exchange will be further elaborated under the Chapter ‘[Extensions](#)’.

### 3.2.2 Flattened table in Excel format

A flattened table in excel format is available [here](#). The flattened table can be used to make a mapping from an existing data structure to the generic data structure for farm data collection, storage and exchange.

## 3.3 Schema details in JSON

### 3.3.1 Farmer group

**Definition** Loosely defined group of farmers. Farmers can be member of a cooperative or union, organized by a company as contract farmers or as part of a support program, etc.

**Data attributes** Many attributes can be assigned to a group, also depending on the nature of the group. Some examples of common attributes are provided, like the area of operation, the valid certifications from standard organisations, total amount of certified produce in stock. Details will not be specified in depth in this version of the reference framework for first mile farm data collection, storage and exchange. The entity is presented here to demonstrate a coherent framework which is ready to be expanded or customized depending on the requirements.

As explained in the section ‘[Uniquely identifying data elements](#)’ {GroupRecordGlobalID, GroupInternalID, Group-GeoID} provides the means to: 1 globally uniquely identify all the records of farmer groups, 2 to provide each farmer group with a unique internal ID which allows to link other data entities to this farmer group and; 3 to have a clue to uniquely identify this farmer group in the field or to clean the dataset when combined with datasets from other sources.

**Datastructure** The data structure is visualised in the table below. All required data attributes are indicated in bold.

- By clicking on the blue table title -> all data attributes become visible,
- By clicking on the data entity buttons -> all data attributes become visible of that data entity and,
- By clicking on the { } symbols -> the JSON becomes visible

### 3.3.2 The farmer

**Definition** The farmer is the person that manages one or more farms, possibly helped by farm workers. The farmer takes the major management decisions even when the decision is to do contract farming where the farm practice is often prescribed in detail by an external actor. In this model the farmer is not necessarily the owner of the farm. A farmer can manage someone else’s farm. Nor does he or she need to be the owner of the assets composing the farm, e.g a sharecropper is seen as a farmer in this model. For a more detailed explanation on how ownership plays a role in modelling different farm types, see also the conceptual model in ‘[the approach](#)’ section.

**Data attributes** groupInternalIDs is an array of database specific identification numbers linking a farmer to 0, 1 or more groups.

As explained in the section ‘[Uniquely identifying data elements](#)’ {GroupRecordGlobalID, GroupInternalID, Group-GeoID} provides the means to: 1 globally uniquely identify all the records of farmer, 2 to provide each farmer with a unique internal ID which allows to link other data entities to this farmer and; 3 to have a clue to uniquely identify this farmer in the field or to clean the dataset when combined with datasets from other sources.

**Datastructure** Required data attributes are indicated by grey shaded fields in the table below.

- By clicking on the blue table title -> all data attributes become visible,
- By clicking on the data entity buttons -> all data attributes become visible of that data entity and,
- By clicking on the { } symbols -> the JSON becomes visible.

### 3.3.3 The farm

**Definition** A farm is described as a collection of assets managed as 1 entity with the primarily objective to raise living organisms for food or raw materials. A farm often consists of fields, buildings and livestock and is run by a farm manager and the farm owner has the rights on the produce. The farm manager and the farm owner are often united in 1 person: the farmer. A farm manager is often supported by farm workers to carry out the activities.

**Data attributes** In the data model the farm is modelled as a placeholder for foreign keys uniting all relevant data entities based on their keys together as 1 farm, including the farmer, the farm owner, the plots of the farm, the herds and the workers etc.

Additional data entities describing, the farm as a whole can be added to this data entity, like the farm boundary, farm area, or aggregated numbers like total amount of produce. Plot specific attributes should be added to the plot like the produce from a specific plot or a plot boundary.

As explained in section ‘[Uniquely identifying data elements](#)’ {GroupRecordGlobalID, GroupInternalID, Group-GeoID} provides the means to: 1 globally uniquely identify all the records of farm, 2 to provide each farm with a unique internal ID which allows to link other data entities to this farmer group and; 3 to have a clue to uniquely identify this farm in the field or to clean the dataset when combined with datasets from other sources.

**Datastructure** Required data attributes are indicated by grey shaded fields in the table below.

- By clicking on the blue table title -> all data attributes become visible,
- By clicking on the data entity buttons -> all data attributes become visible of that data entity and,
- By clicking on the { } symbols -> the JSON becomes visible

### 3.3.4 The plot

**Definition** A plot is an area of land, enclosed or otherwise, used for agricultural purposes such as cultivating crops or for livestock.

**Data attributes** The data entity plot describes the main and general characteristics of a plot: including location and geometry, but also agronomically. Examples are:

- seed variety (name/vendor/date/quantity in kg/ha)
- yields (kg/ha)
- pesticide use (product/trade name/vendor/date/ quantity in kg/ha)
- fertilizer use (product/trade name/vendor/date/quantity in kg/ha) and application method
- measured or calculated water use (date/quantity in kg of harvested paddy/litre water input)
- timing of the different farm activities

As explained in section ‘*Uniquely identifying data elements*’ {GroupRecordGlobalID, GroupInternalID, Group-GeoID} provides the means to: 1 globally uniquely identify all the records of plots, 2 to provide each plot with a unique internal ID which allows to link other data entities to this plot and; 3 to have a clue to uniquely identify this plot in the field or to clean the dataset when combined with datasets from other sources.

The model can be expanded by creating rotation specific, crop specific data entities or data entities describing the agronomic conditions like soil type.

#### Additional standardization resources

- [World reference base for soil resources 2014](#)

**Datastructure** Required data attributes are indicated by grey shaded fields in the table below.

- By clicking on the blue table title -> all data attributes become visible,
- By clicking on the data entity buttons -> all data attributes become visible of that data entity and,
- By clicking on the { } symbols -> the JSON becomes visible

### 3.3.5 The observation

**Definition** Data obtained from an assessment in a plot, generally done at a specific location, in a transect or in an area. This data entity is not meant to describe the plot as a whole. Based on observations estimates are made for the characteristics of the plot as a whole. The observations are records of the observation data entity, where as the estimation for the plot as a whole based on the observations should be added as a data attribute of the plot data entity. For example in a grass plot the biomass is measured at 3 points. This can be added to the data set as 3 observation records, the average biomass calculated for the plot can be added to the plot data entity.

**Data attributes** Typical observations are:

- crop height,
- number of branches,
- height of Jorquette,

- percentage of trees show old wilted,
- number of dead or mummified pods,
- number of dead branches,
- number of damaged or diseased pods,
- number of epiphytes
- number of ant nests/tunnels nests,
- etc.

As explained in section the ‘*Uniquely identifying data elements*’ {GroupRecordGlobalID, GroupInternalID, Group-GeoID} provides the means to: 1 globally uniquely identify all the records of observations, 2 to provide each observation with a unique internal ID which allows to link other data entities to this observation and; 3 to have a clue to uniquely identify this observation in the field or to clean the dataset when combined with datasets from other sources.

### Additional standardization resources

- [ISRIC - WoSIS exchange requirements](#)

**Datastructure** Required data attributes are indicated by grey shaded fields in the table below.

- By clicking on the blue table title -> all data attributes become visible,
- By clicking on the data entity buttons -> all data attributes become visible of that data entity,
- By clicking on the { } symbols -> the JSON becomes visible.



The reference framework is designed in such way that new data attributes or even new data entities for different purposes can be added to the model easily. This is illustrated below for Cocoa action program and MARS Adoption Observations. The data formats are available together with the official protocols on how to collect the data in the field.

By adding more and more extensions to the reference framework, a repository emerges of ‘proven’ data formats and data collection protocols. Organisations can benefit from this repository harvesting the formats and data collection protocols they need for their own data management. Having different data formats and data collection methodologies in 1 repository will also facilitate further standardization and interoperability discussions. To add to the reference framework please contact [andre.jellema@data-impact.com](mailto:andre.jellema@data-impact.com).

### 4.1 MARS adoption observations

This is an unofficial extension schema of the reference framework for first mile farm data collection, storage and exchange, intended as an example for the MARS Adoption Observations. For further information on MARS Adoption Observations see [the ppt](#)

**Added data entities and data attributes** In the MARS adoption observations framework questions are being asked about the farm. However when carefully read all data needs seem to address generalizations of the crop at plot level. Hence the Mars adoption observations have been implemented as plot level data entities.

#### **Prescribed methodologies to measure the data elements**

Description of the methodologies go [here](#).

### 4.2 CocoaAction

This is an unofficial extension schema on the reference framework for first mile farm data collection, storage and exchange, intended as an example. This schema is building on the standardization discussions in the cocoa valuechain. For further information on CocoaAction see [the website](#)

**Added data entities and data attributes**

**Prescribed methodologies to measure the data elements**

Find the Cocoaaction M&E guide [here](#) with the descriptions on how to measure the different dataelements.

**Mapping of all CocoaAction data on the First-Mile Farm Data Schema**

---

## 5 Uniquely identifying data elements

---

### Key message

**In the reference framework a system of 3 identifiers is proposed to uniquely identify objects and model the relationships between these objects.** These Identifiers are called: 1 the InternalID, 2 the GeoID and 3 the GlobalID. The system allows the exchange and merging of data that is structured in a similar way.

1. **The GlobalID** is a globally unique identifier for first mile Farm data record.
2. **The InternalID** provides a means internally structure the data set by uniquely identifying the objects in the dataset and relations between them.
3. **The GeoID** provides an imperfect means to uniquely identify the object in the field and to clean the dataset when it is combined with datasets from other sources.

The Internal database structure is provided by the InternalID. The GeoID and the GlobalID can be used to maintain this structure while exchanging data between organisations.

### 5.1 Introduction

In an ideal world every farmer, farm, plot or farmers group would be globally uniquely identified by some formal or informal system and these identification numbers would be generally known and used by different organisations. Organisations could make data recordings about farmer, farm, plot or farmer groups and could easily combine these recordings with other recordings trusting these globally unique IDs to sort the data and make analyses using all records combined. However currently this is an utopia, because there is no perfect system uniquely identifying all farmer, farm, plot or farmers groups globally (see also the sections below).

Instead the reference framework uses 3 identifiers to build and internal structure and maintain a local data set of first mile farm data: 1 the internal ID, 2 the GlobalID and 3 the GeoID.

## 5.2 The InternalID

The InternalID is the ID used to uniquely identify Data Entities in a data set in a local data base and to structure the data using these internalIDs as Keys to link the farmers to farmer groups, farms to farmers, plots to farms and plot observations to plots.

When the data set is merged with another first mile farm data set the internal ID needs to be updated to represent the new structure of the data sets. In principle there are **3 problems that need to be solved**.

1. **If 2 datasets are merged and they may contain information about the same farmers** these farmers should get the same InternalIDs and all related Keys inside the combined dataset need to be updated.
2. **If 2 different farmers have the same InternalID before the merge**, 1 of the 2 farmers need to get a new and unique InternalID during the merger and all Keys need to be updated accordingly.
3. **If dataset a is shared with organisation b and with organisation c and next these organisations merge their datasets** the newly derived data set will contain all records of dataset a twice. These doubles need to be removed from the database.

**Datastructure** The InternalID is a string with a Universally unique identifier used in many databases to identify dataelements.

## 5.3 The GeoID

**To clean the datasets from problem 1 and 2 the GeoID is introduced.** The GeoID is the location of an object in WGS84 coordinates and a description of where the geolocation is measured. When the WGS84 TRF (*see also the section on spatial data acquisition and standards*) is viewed as an register using 2 ID numbers, WGS84 can be considered as the only registration system that is consistently used globally to uniquely “identify” objects over the past 20 year. However the location of an object may have different interpretations between different data collectors, therefor a description of where the geolocation is measured is made part of the GeoID to ease the cleaning process. One can imagine that the cleaning based on geolocation will be made with much more confidence when all GeoIDs are taken at the doorstep of a farm and not at the location where the interview is taken. However sometimes the GeoID can not be taken at the doorstep, therefor different options are provided with decreasing accuracy:

1. Location of main entrance of the object or the front door of the main office from where operations take place;
2. Centerpoint of the object or centerpoint of area of operation of the object;
3. Location on the object or in the area of operation;
4. Location near the object or near the area of operation of the object;
5. Location where the interview has taken place

It is assumed that even the location of an interview is relative in the proximity of the place where a farmer has is activities. It may be the only information available and is still informative on the area where the farm is located. Therefore it is recommended to add additional identifying information if available to facilitate the cleaning process, like:

- the legal or official name
- the legal registration number (and what kind)
- the Address
- etc. (*see also the section below on alternative ways of identification for more suggestions*)

**The GeoID also provides a means to verify an identities of objects in the field.**

**Datastructure** The supporting data structure is visualised in the table below. All required data attributes are indicated in bold.

- By clicking on the blue table title -> all data attributes become visible,
- By clicking on the data entity buttons -> all data attributes become visible of that data entity and,
- By clicking on the { } symbols -> the JSON becomes visible

## 5.4 GlobalID

**To solve problem 3 the GlobalID is used.** The GlobalID is considered an indivisible part of the data record and attached to it at the time of recording. The GlobalID consists of:

- The original internal ID used by the data collection organisation to identify that farmer, farm, plot, farmers group etc; When objects are uniquely identified only at project level, the project ID needs to be added to the object ID for example (Internal Project ID + Internal Object ID).
- The latitude and longitude coordinates of the main entrance of the headquarters of the organisation that collected the data;
- Time-Date stamp of the recording.

The GlobalID always needs to stay attached to the data record and should not be changed when exporting, merging or updating datasets. The GlobalID also provides a means to go back to the source of the data.

In principle the reference framework is building a pool of globally uniquely identified data records. If the data model and the 3 identifiers are used consistently by its user group, the combined dataset of uniquely identified records can be considered as the global pool of first mile farm data allowing for global data exchange and data analysis.

**Datastructure** The supporting data structure is visualised in the table below. All required data attributes are indicated in bold,

- By clicking on the blue table title -> all data attributes become visible,
- By clicking on the data entity buttons -> all data attributes become visible of that data entity and,
- By clicking on the { } symbols -> the JSON becomes visible

## 5.5 Alternative ways of identification

All of these data could be added to the records to facilitate the identification in the field and the cleaning of datasets.

- **Using a national register** The most straightforward way to uniquely identify a farmer, a farm or a plot of land would be to use an official national registration number, such as the tax number, a social security number, a civil service number, a company register number or a land registration number. However, in many countries such systems do not exist or are only partially complete and therefore no solution that can be applied globally. Many of these data sources are also privacy sensitive. <http://org-id.guide/> provides an overview of organisational registers in different countries.
- **Using a Cooperative register** Many farmers work for a cooperative. In order to function well, the cooperative needs to register all its members to administer things like: the farmer's input needs, the amount of produce delivered to the farmer, debit and credits, etc. The reliability of the data in the administration of a cooperative varies to a great extent. Moreover not all farmers are connected to a cooperative, making it impossible to create a globally unique identifier for all.

- **Using the Blue Number** The [Blue Number Initiative](#), led by International Trade Center (ITC), the UN and GS1, aims to make food and agriculture supply chains and systems more sustainable, contributing to the United Nations Global Goals for Sustainable Development. The Blue Number acts as a unique identifier for use by anyone involved in the food chain from farmers, producers, distributors and vendors to consumers. Currently there are about 2000 people registered, mainly in Singapore.
- **Building an enumeration of real world characteristics.** Uniquely identifying a farmer by his or her name works pretty well in small groups. However, when the database expands many doubles will occur. As strategy to stratification with real world characteristics can be created in the database to maintain the small groups with mostly unique names. An examples of such a stratification is to add the names of the lowest and second lowest administrative zones to the name the farmer.
- **Universally unique identifier** Many technology companies use a universally unique identifier (UUID) to identify farmers and other objects in their databases. A UUID is a 128-bit number which is generated to identify information in computer systems. When generated according to the standard methods, UUIDs are for practical purposes unique, without depending for their uniqueness on a central registration authority or coordination between the parties generating them, unlike most other numbering schemes. While the probability that a UUID will be duplicated is not zero, it is so close to zero as to be negligible. However these UUID are generally applied within the context of one organisation or even one project. This means that the same farmer will have two different identifiers between 2 organisations or even 2 projects. This means that when data sources from a different origin are being combined for re-use, one farmer can appear in 2 or more records with different ID numbers.

---

## 6 Spatial data acquisition and standards

---

### Key messages

Spatial data collection is always in reference to some theoretical model of the earth. The recommended ‘*model earth*’ or Terrestrial Reference Framework (TRF) to locate a point on the earth’s surface is GWS84 using decimal degrees Latitude and Longitude coordinates with at least four decimal places (d.dddd), for example Latitude: 52.0525, Longitude 5.2667. In this way locations are indicated with an accuracy of 10m.

**Using a Navigational Satellite System (NSS)** Modern smartphones have enough accuracy to map geolocations in the field using the internal receiver for NSS signals. Depending on the conditions accuracies between 2 and 9m can be reached.

The NAVSTAR GPS system is the standard system currently in use on most mobile devices, however some devices are capable of combining GPS with other navigation systems, providing a higher accuracy and shorter search time for satellites connections.

Make sure that the accuracy while measuring is < 10m. Automated averaging of multiple measurements of the same location is a practical solution to reduce error towards 1m.

To measure the boundary of an area (a polygon) only the corners should be measured manually minimizing the error of the area measured. On the contrary a strategy of measuring multiple points along the edge of an area or using an automated tracking function in your device reduces the accuracy.

**Using screen digitization** Locations and areas can also be mapped from a screen. Screen digitization is always in relation to another data source (aerial photograph, satellite image or GIS file). There are a number of points to take into account:

- The TRF of the reference source and the TRF that is used for the digitization. Local datasets used are possibly based on a different TRF. The combination of data sources with different TRF’s or other data handling using the wrong TRF leads to errors in the results. More advanced desktop GIS, like QGIS or ARCGIS are capable of making data transformations from one TRF to another TRF but need to be operated by a specialist.
- The accuracy of the reference data set is determining the accuracy of the end result. Field boundaries or boundaries between farms may not be clear on an aerial photograph for small holder farms. Boundaries may not be visible or appear fuzzy. It is best to check the quality of a third party data set by overlaying it with actual field measurements or other data with known accuracy.

Different approaches to digitize data from screen are described in this section.

### 6.1 Introduction

The collection of spatial information of objects, like the geolocation or the shape of objects, is of growing importance for first mile data collection. Spatial information helps :

1. to identify objects (see the section on *‘Uniquely identifying data elements’*,
2. to combine different sources of information, e.g. to identify all farms with a particular soil type using a soil map,
3. to determine spatial relations, e.g. to select all farms with a market within 10 km and to determine spatial characteristics of an object e.g. the area of a field.

**The geolocation** is a description of the location of an object at the earth’s surface in coordinates relative to a mathematically defined **Terrestrial Reference Frame (TRF)**. The standard TRF recommended in this document is GWS84 using decimal degrees Latitude and Longitude coordinates with at least four decimal places (d.dddd), for example Latitude: 52.0525, Longitude 5.2667. Four decimal places indicate an accuracy of about 10 meters and five decimal degrees indicate an accuracy of about 1 meter. Points in the Southern hemisphere have negative latitudes; points in the western hemisphere (the Americas) have negative longitudes. If the sign is positive (northern hemisphere latitudes and eastern hemisphere longitudes), it is not necessary to include a “+” sign. The **WGS84 TRF** is the first globally applicable TRF and is currently the default for world standards and for most applications and devices including the **GeoJSON** and the NAVSTAR Global Positioning System or **GPS**. When using spatial data, especially when using third party georeferenced data, it is essential to be aware of the coordinate system used to describe the geolocations. Combining 2 spatial datasets based on a different TRF will result in errors if not carried out properly. More background on fundamentals from geodesy and cartography needed for georeferencing can be found in [the section on the complexity behind geodata](#).

Nowadays spatial information is often collected and stored digitally using an [object model of points, lines and polygon or collections of these](#). A Point type object is described by a single coordinate pair. A line string is a one-dimensional object represented by a sequence of points and the line segments connecting them. A Polygon is a two-dimensional surface stored as a sequence of points defining an exterior bounding ring of at least 3 points and zero or more interior rings to represent holes in the polygon if needed. Spatial information about objects can be collected digitally in the field using a Navigational Satellite System (See [section 6.2](#)) or using a geographical information system(See [section 6.3](#)).

### 6.2 Using a Navigational Satellite System to collect spatial data in the field.

A **Navigational Satellite System (NSS)** is a constellation of satellites with accurately known positions or orbits that transmits coded radio signals. These coded radio signals are used by specialized electronic devices to calculate the location on the earth’s surface relative to the satellite constellation. Currently, the American **NAVSTAR Global Positioning System or GPS** is the most used system. However alternative systems exist, or become operational soon in the future (see also the below table below). Nowadays some receivers can already use the data from these alternative systems, or combine the signals of several systems. Combining the data from 2 or more satellite systems increases the accuracy of the measurement and decreases the search time for satellites data connections. Measurement with a NSS has an error. This error is depending on the system used, the openness of the sky above the location device (no buildings or trees), the number of satellites above the horizon at the moment of the measurement and the actual position of these satellites in the sky. Under optimal conditions the error varies between 2 and 10 meters.



### 6.2.1 Collecting a point location

*The following procedure is derived from the SAN guidelines for certificate mapping produced by the Rainforest Alliance.*

There are many mobile apps that turn an Android or iOS smartphone into a mapping tool; the following are just two examples that are free, easy to use, and work offline. These programs should be configured in the settings to report in decimal degrees. It is best to report location points by saving “waypoints” and downloading these onto your computer; this avoids the possibility of transcription errors. It is also possible to read the latitude and longitude coordinates directly off the app display and record this information in a notebook or data form. To increase the accuracy of your measurement towards 1m, multiple measurements can be taken at the same location and then averaged. Some GPS systems have an automated option for this.

**Using GPS Essentials** This app provides a suite of location tools, including a tool for collecting points (waypoints). To collect a location point with GPS Essentials, click on the waypoints icon and the orange “plus” button in the bottom right corner (see screenshots below). Next, wait for a moment for your phone to pick up satellites and improve accuracy. The accuracy of your GPS should be 10m or less. Now, enter the name and description of the waypoint and click “create”; you can now read the latitude and longitude location of your point. If you like, you can save your point and export it to your computer. To do this, open the “Waypoints” button, click the three dots in the upper right corner for the pull-down menu and select “Export” to email the data to yourself as a .kml file.

**Using My Maps** Google My Maps is a mobile app and also a web-based application that runs on your desktop. The My Maps mobile app is similar to the Google Maps app that comes with many phones; however, it has additional functionality and must be downloaded and installed as a separate application. Once you have installed the app on your device and logged in with your Google account, you can create your own maps and add points. When you return from the field and connect with the Internet, these mobile maps synchronize with your desktop web-based My Maps program, where the data can be exported as a .kml file.

When in the field, launch the Google My Maps mobile app. To add a waypoint, click the blue “+” sign in the lower right corner (see figure x), click on “Add a new point”, select the location, select the layer, and give the point a name and description. You also may be able to add points by holding down at a point in the screen for a few seconds; then you will be prompted to name and save this point. You can now read the latitude and longitude location of your point. If you like, you can also now view the point in Google Maps – My Maps, since the point will automatically sync through your Google account.

### 6.2.2. Collecting a polygon in the field.

A polygon is a 2 dimensional feature defining the border of a surface area with a series of points. In the polygon data model the points are interconnected by line segments in the same order as they are recorded. A polygon can readily be used to map a field or a waterbody. The best way to map a field or a waterbody is to walk around the object in the field and to only measure the corners of the fields manually as a waypoint. It is not recommended to use an automated trace function to measure the boundary or to take multiple points along the boundary of a field, this decreases the accuracy or the measurement:

- First of all, when a trace function is used, any deviation from the boundary, caused for example because part of the boundary is not walkable will be recorded. This will result in an oddly shaped boundary and higher inaccuracies for area calculations.
- Secondly, when measuring distances and areas, the relative error in the measurement decreases with the distance between the measurements. For example if 2 points are located 20m apart and the GPS has an accuracy of  $\pm 5\text{m}$ , the relative error in the measurement is  $\pm 50\%$ . However if the distance between 2 points is 200m with the same accuracy, the relative error decreases to  $\pm 5\%$  only. The same mechanism applies to area calculations. In conclusion: to increase the accuracy of distance and area calculations, the distance between the point measurements

need to be maximized. This happens when only the corner points of polygons are being measured and not the inbetween points along the edges.

## 6.3 Using a Geographical Information System (GIS) to collect spatial data.

Another approach is to digitize areas and locations from a screen using a geographical information system or GIS. A IS can be desktop or webbased. The 2 main desktop applications are the open sources [QGIS](#) and [ArcGIS](#). Both application has similar functionalities and need a specialist to operate. The community [Geo for All](#) has the mission to make geoanalysis accessible for all, providing freely accessible training programs on the use of QGIS. Beside desktop GIS systems, webbased GIS systems are available. Most of these web apps provide visualization and digitization functionalities. Examples are [Google Earth](#) and [Google My Maps](#) and are available free of charge.

Screen digitization is always in relation to another data source (aerial photograph, satellite image or GIS file). There are a number of points to take into account:

- The TRF of the reference source and the TRF that is used for the digitization. Local datasets used are possibly to based on a different RTF. The combination of data sources with different TRF's or other data handling using the wrong TRF leads to errors in the results. More advanced desktop GIS, like QGIS or ARCGIS are capable of making data transformations from one TRF to another TRF but need to be operated by a specialist.
- The accuracy of the reference data set is determining the accuracy of the end result. Field boundaries or boundaries between farms may not be clear on an aerial photograph for small holder farms. Boundaries maybe not visible or appear fuzzy. It is best to check the quality of a third party data set by overlaying it with actual field measurements or other data with known accuracy.

*The following procedure is derived from the SAN guidelines for certificate mapping produced by the Rainforest Alliance.*

### 6.3.1 Point locations

If the location of interest can be identified in a web-map or via satellite view, then it may be easy to obtain the GPS coordinates from the screen. In Google Earth, just place the cursor at the desired location and read off the coordinate values from the “Status bar” at the bottom of the screen. In Google Maps, simply create a point or marker at the desired location and then click on that marker to display the properties, which will show the GPS coordinates (usually at the bottom left of the pop-up). When reporting these points, make sure that they are in Decimal Degrees.

### 6.3.2 The digitization of polygons

*Using Google Earth* Google Earth Pro is a program that you can download for free and install on your computer ([download here](#)). After installing the program, select from the main menu Tools -> Options and configure: show Lat/Long in Decimal Degrees (3D View tab).

The Google Earth screen is divided into 3 panels on the left-hand side of the screen and one larger panel with the “Map Viewing Area” in the center (see screenshot below). For the purpose of mapping farm polygons, you will mainly utilize the “Map Viewing Area” and the “Places panel”, which is where your data can be referenced and organized. The content layers under the folder “My Places” are automatically copied to your hard drive when you exit the program and re-loaded the next time it is launched.

If you have reference points collected in the field in .kml or .kmz format. They can simply be loaded into Google Earth by double-clicking on the .kml or .kmz file. Alternatively, you can also open kml/kmz, gpx and many other types of spatial data by clicking on File -> Open. Data imported into Google Earth are

stored in the “Temporary Places” folder of the “Places panel.” You must remember to move the files up to one of your folders so that they are not lost when you exit the program.

When time-stamped GPS data are imported and displayed in Google Earth, the “Time slider” is automatically displayed at the top of the screen (figure 5). This slider has “Range Makers” that control the time and date range of the tracks and points that are displayed. This is important because sometimes you might think your data has disappeared, when it is only out of the range specified in the time slider. To view all the data, move the range markers to the far right and left sides to define a wider date range.

To draw a polygon, follow these steps (figure 6):

1. Select the *Add polygon* tool.
2. Click on the map at the location of the polygon corners (vertices), going around the entire edge of the polygon to define its shape.
3. When done, give the polygon a name in the “Name” field and add any additional details in the “Description field.”
4. Click on the “Style, Color” tab to define how the polygon is displayed.
5. Click on *OK* to save the polygon.

Once you have created your polygon you should see it in the left Places panel. You can edit both the vertices and the properties by right-clicking on the item in the Places panel and selecting properties. To save the polygon(s) as a kml/kmz file, in the Place panel move all the polygons you want to save into one folder, right-click on the folder, select “Save Place As” and enter the name and location of the output file.

**Using polygons with Google My Maps** Google My Maps is an extension of Google Maps. It can be used as part of Google Maps on any internet browser (after you sign in using your Google account); on a smartphone it must be downloaded as a separate app. All maps and data created in either form can be displayed and edited in the other. Since maps are web-based, they can be shared and edited by multiple users. To use [Google Maps](#), open the website and sign-in with your Google credentials. This will take you to the >My Maps home screen, where you can create a new map (figure 8). For a tutorial on how to create a Google My Maps [clickhere](#).

Click on the “Create a New Map” button in the upper left corner to initiate your new map. Now you are ready to start using your map. Give the map a name, zoom-in to your area of interest and select the basemap from the various options (map, terrain, satellite, atlas and more).

**Upload data into Google My Maps** As Google My Maps is entirely web-based, whatever data you collect in the field on the My Maps app (waypoints, lines) will automatically sync into the desktop map. It is also possible to import spatial data stored as .kml, .kmz, .cvs or .gpx files. To do this, first click on *Add layer* and then click on the *Import* button that appears under the new layer. A window is displayed for you to upload or drag / drop files to be imported.

**Create and edit polygons in Google My Maps** Use the tools in the upper center of the map to create points, lines and polygons. To create a polygon, use the line tool and go around the perimeter of the desired polygon area while clicking on the map to create a vertex at every corner. To complete the polygon click again on the first point. If you have collected reference points from the field and imported them to Google My Maps, these can help indicate where to draw the polygon vertices. When you are finished, enter the polygon name and description, and save.

All features created can be symbolized by selecting the line thickness, icon type, and color. Note that when you create a point, the latitude / longitude coordinates are displayed at the bottom of the point’s information window. Maps are automatically saved as you work.

To export the map, click on the three dots in the top right of the map box, select “export to kml” and select whether you want to export all the map data, or just the data in a specific layer. Note that Google My Maps has limited ability to move and organize features into layers, so for complex sites, it might be

necessary to export the map into a more full-featured program, such as Google Earth, to organization the multiple polygons.

## 6.4 Some more background on the complexity behind geodata

When working with spatial data it is important to be aware of the complexity behind it. Locations in space are generally described in mathematics, as points in a rectangular coordinate system consisting of three axis (x,y,z). However, when we are thinking and talking about the world around us in daily live, we are actually mentally flattening the curved surface of the earth in the 2 dimensions (x,y) of a flat plane and define height (z) as the elevation of the surface above sea level. This section explains some fundamentals from geodesy and cartography to get an understanding of this complexity in order to avoid mistakes during spatial data handling because of misinterpretations.

The description of the earth surface in coordinates is based on mathematical models called **Terrestrial Reference Frames (TRF)**. A modern TRF is defined by a *Datum* and a *Geoid model* of the earth.

*source=<http://www.dauntless-soft.com> source=<http://www.asu.cas.cz>)*

A **Datum** is a numerical description of the earth surface as an ellipsoid. This ellipsoid is the used to pinpoint locations on the earth's surface in latitude and longitude coordinates. **Latitude** is a geographic coordinate that specifies the north-south position of a point on the Earth's surface. Latitude is an angle which ranges from 0° at the Equator to + 90° at the North Pole and -90° at the South pole. Lines of constant latitude are called parallels and run east-west as circles parallel to the equator. **Longitude** is a geographic coordinate that specifies the east-west position of a point on the Earth's surface. Lines of constant longitude are lines running from the North Pole to the South Pole called meridians. The longitude of a location is measured as the angle east or west from a reference Meridian, ranging from 0° at the reference to +180° eastward and 180° westward. Different Datums are being used to describe different parts of the world, resulting into different locations for the same location when using the same coordinates in different reference frames. The distance between 2 points with the same coordinates using different datums may go up to several kilometers. Height in daily live is measured as the elevation of a point above sea level. This measurement is challenging when it needs to take place in landlocked country in the middle of a continent. As a starting point in a modern TRF uses an Earth Gravitational Model (EGM) to model the mean sea level for all water surfaces AND across all land surfaces. This model of the mean sea level across the globe is called a **Geoid**. Because the gravitaional field varies across the globe is the resulting model irregularly shaped. The Geoid's of modern TRFs match each other **within the centimeter meter accuracy**, taking **EGM** as a starting point for the model.

Globally the **WGS84 TRF** developed by the US Department of Defense is the most widely used TRF for many years now as a result of the rise of the usage of **Global Positioning Systems (GPS)** for field work, mapping and navigation. This drastically simplified the exchange of spatial information and logically the WGS84 is also the standard recommended in this document. However there are still 2 points to be aware of:

- The WGS84 reference system has been updated several times in the past and will be updated in the future. For the mapping purposes as discussed in this document **the WGS84 can be considered stable since 1996**. Before this date the WGS84 TRF was using an ellipsoid rather than a geoid as a reference to express the height above the earth's system. The difference in height between the WGS84 TRF and the height value based on the modern WGS84 is between 105 m and +85 m. **The GeoidEval utility** can be used to calculate the height difference between the old model and the new model at every location on earth. In addition paper maps or digital data coming from external sources may be using different datums. There is a whole history in mapping and geodesy finding the most accurate representation of the earth in different regions. The site **Coordinate Systems Worldwide** provides overview of all TRFs in use and the transformations between.

In practice when we are working with spatial data (e.g measuring distances, working on a paper map, a computer screen and even when using the GPS map display) we are working with spatial data presented on a flat surface not on a globe model. The mathematical transformation from a ellipsoid representation of x,y coordinates on a flat surface is called a map projection. To illustrate how different concepts work common projections are visualized by imaging the globe as a wire-frame model with a light source in the center, the shadows created beyond the sphere can be "projected" onto a flat surface.

*source=Dylan Prentiss, Department of Geography, University of California, Santa Barbara.*

Every type of projection is an imperfect representation of your data and will be distorted in some way. Some methodologies are specifically capable of preserving area relations, shapes objects or the distance between points, but none will preserve all. There are endless ways to make map projections depending on the region and the purpose of the map and this is important to realize when using external data sources. The Universal Transverse Mercator projection is popular world wide applicable map projection system. <http://epsg.io/> also provides a comprehensive database and access to all coordinate systems worldwide, including the map projection and the TRF and mathematical transformations between them. The data based can easily be searched for your region by clicking on a map.

Popular applications, like google maps, bing maps, openstreet maps etc. are using the webmercator projection. Despite the popularity of these applications it has taken a long time before the webmercator projection was accepted by the Geodetic community as an appropriate map projection system. The key problem of this system is that the ellipsoid model of the earth to project lat long coordinates on a plane is replaced by a spherical model to speed up calculations and internet response times. This results in systematic errors up to 40 km's when compared to a spheroid projection based on WGS84. In addition the 'primitive' mercator projection, dated from the middle ages, heavily distorts distances and areas at latitudes further from the equator, for example Greenland is projected with similar dimensions as latin America. The key advantage of the mercator is that it allows for effortless zooming in a squared world and angles between points on the map are preserved. In online applications as google maps, ArcGIS Online or bing maps the errors in distance and area calculations are generally fixed making calculations in an alternative TRF. However the National Geospatial-intelligence Agency still warns for the incompatibility with and non-compliance of this type of applications to the WGS 84 TRF. This ESRI website illustrates the differences that are produced by different map projections stressing the need to be aware of the TRF and map projection that is used while working spatial data, especially while calculating distance and surfaces.