

Guidelines for the Collection and Management of Location Data in the ISEAL Certification Atlas

David Hughell, Elizabeth Kennedy and Adam Romo

Draft: 17/October/2017

TABLE OF CONTENTS

1	What is the Certification Atlas?	2
1.1	Proposed objectives of CA.....	2
1.2	Purpose of guidelines	2
1.3	Terms defined	3
2	Structure and content of certificate location data	3
2.1	Questions about the structure of the certificate.....	3
2.2	Types of location data	4
2.2.1	Point	4
2.2.2	Polygon.....	4
2.3	Permanent and unique ID codes for locations	5
2.3.1	Bluenumbers.....	6
2.3.2	GS1 barcodes.....	6
2.3.3	Geography-based code	6
2.4	Location attribute data.....	7
3	Collection of location data.....	8
3.1	Field data collection using GPS.....	8
3.2	Turn your smartphone into a GPS	9
3.2.1	GPS Essentials (http://www.gpsessentials.com)	9
3.2.2	Google My Maps (https://www.google.com/maps/about/mymaps/).....	9
3.3	Collection of polygon data in the field.....	10
3.4	Digitizing of location data using interactive maps or GIS	11
4	Management and processing of spatial data.....	11
4.1	Tabular and point data	11
4.2	Geospatial polygon data.....	11
4.3	GFW Pro	12
4.4	Certificate Atlas platform	12
5	Data disclosure, and security.....	13
5.1	Sharing aggregate data.....	13
6	Conclusion and areas of discussion	13

Executive Summary

These guidelines have been prepared to serve as the basis for discussion on standardized procedures and protocols for the collection and management of certificate location data with the goal that these data will eventually become part of an ISEA member Certification Atlas (CA), either as individual site locations or in aggregate form. Included here is a discussion of data formats, field collection methods, data processing procedures and policies for data sharing. ISEAL's approach is to be "software agnostic," however, several programs are recommended that members can use. As such, WRI's GFW Pro program is presented as a tool that members can use to upload and manage certificate data, and carry-out pre-established land cover and risk analyses.

1 What is the Certification Atlas?

The ultimate aim for an ISEAL Certification Atlas (CA) is to be a public resource that displays where on the planet operations that have been certified against sustainability standards of ISEAL members are located. These locations represent where we would expect to find the direct outcomes created by the extraction, production, fabrication and/or provision of certified goods and services. Therefore, the CA will contribute to providing transparency in the certification process (ISEAL credibility principle 2) and will establish the starting point for traceability (ISEAL credibility principle 9). The CA can be thought of as the intersection of the certificate (the basic unit of evaluation tracked in the certification process) and the fixed geographic location where the certificate holder activities and, consequently, their impact, occurs.

1.1 Proposed objectives of CA

ISEAL is currently undertaking a scoping study to create a unifying vision and strategy for its members and partners to implement a CA. Some proposed objectives for a CA include:

- Provide a one-stop, credibly verified resource for information on the extent of certified areas across all ISEAL member schemes globally.
- Facilitate the validation of compliance with geographical sustainability criteria, such as the maintenance and preservation of critical habitats.
- Make available a data layer that can be integrated with other geospatial information such as land use, areas of high conservation value, and social and environmental pressures, to create a basis for more robust and targeted interventions, or conducting collective and tailored analyses of risk.
- Promote landscape and jurisdictional certification efforts, which encompass multiple goods and services, multiple certification schemes and multiple-stakeholders.

1.2 Purpose of guidelines

This document provides standardized protocols and recommends procedures to collect, manage and submit data on the location of certified operations for inclusion in a CA. ISEAL's approach to managing location data is to be "software agnostic," which means a specific program is not required for the protocols and data formats to work. An important aspect of being software agnostic is data interoperability, such that data entered into one program can be easily exported and uploaded into another. To be software agnostic we must provide adequate detail so that the users will know how to apply the protocols and data formats to their programs.

Please consider everything presented here as open for discussion and, as such, we look forward to your feedback.

1.3 Terms defined

Attribute	Tabular data that is associated with a specific certificate, site or location.
Certificate	The basic unit tracked, audited and certified. It might consist of one or multiple sites or locations.
Feature	In the context of GIS, it is a specific point, line or polygon that represents something on the Earth.
Geographic Information System (GIS)	A computer program that integrates geospatial and tabular information for analyses and preparing mapping products.
GFW Pro	A web-based software mapping tool developed by WRI for mapping and processing location data using pre-determined formats and analyses.
Latitude/longitude	The x,y coordinates that specify a specific location point on the planet.
Location	A specific geographic place on the planet that can be defined as either a point or an area with boundaries (polygon).
Global Position System (GPS)	A device or embedded chip that reads signals from a constellation of GPS satellites to determine the location on the planet.
Point	A geographic feature represented by a pair of latitude/longitude coordinates.
Polygon	A geographic feature that defines an area with boundaries.
Shapefile	A format for storing geospatial data.
Site	A fixed location where a certified product or service is provided. One certificate might consist of one or many sites.

2 Structure and content of certificate location data

We start this discussion with the common element of all certification schemes – the certificate. This is what is assigned a unique identification (ID) code, tracked, audited and awarded a certificate for complying with certain sustainability standards so the generated products can be traded as “certified,” ideally within a chain-of-custody traceability system. Depending on the geospatial structure of the certificate, it might occupy one site or multiple sites. These sites are often widely separated geographically. Examples of multiple-site certificates are RA/SAN Group certificates and FSC certificates consisting of more than one noncontiguous forest management units (FMUs). For the CA, our attention is on these single and multiple sites, since they are where sustainability is evaluated, outcomes can be measured and where most impacts occur.

2.1 Questions about the structure of the certificate

Certificate scheme owners should answer several questions that will help define the geographic structure of the certificate:

- For your certification scheme, are certificates defined by established fixed geographic locations? If not, then maybe CA is not relevant to your scheme.
- Do certificate locations occupy an area so small that they can be represented by a point (such as a factory or hotel as opposed to a farm or forest management area)?

- Is each certificate represented by exactly one site, or are there cases where a certificate is represented by multiple sites? The distance between the multiple sites will have a bearing on whether you will want to account each site individually within the CA or not.

The answers to these questions should help decide whether to use the certificate or the site within the certificates (sub-certificate) as the primary unit of analysis within the CA.

2.2 Types of location data

Sites or locations can be represented geographically as points or areas, the latter called “polygons,” depending on availability of data, the area and the scale it is to be displayed.

2.2.1 Point

A point is a dimensionless feature (e.g. it has no area) represented by a pair of latitude/longitude coordinates. Points are easy to manage, since all they require are 2 columns in a data table. A location point can easily be captured with a GPS (ways to do this are discussed later in the Guidelines).

In the context of the CA, we can use points to represent a certificate location that is composed of many sites. It is preferable to represent each site individually. However due to a lack of data sometimes this is not practicable, so we resort to placing one point as the centroid of a cloud of sites (figure 1a).

2.2.2 Polygon

A polygon is a geographic area feature defined by a series of points (or vertices) that are connected to form a ring (and hence enclose an area). Due to their more complex nature, the management of polygons requires special software, such as a GIS or specialized mapping program like Google Earth.

Like points, we can use a polygon to represent one or more contiguous sites (figure 1d), or to represent a landscape of multiple sites (figure 1b). The difference being in the latter, in addition to the certified operations, there are areas within the polygon that are not certified.

In summary, there are 4 ways to represent multi-site certificates geographically, presented here from the most general (least useful) to more specific (and most useful):

- One point representing a certificate with many sites (Figure 1a).
- One polygon representing many sites. (Figure 1b).
- One point for each site (Figure 1c).
- One polygon for each site (Figure 1d).

One intermediate option not mentioned above is to represent each site by a point with a radius. The radius can be a fixed distance or based on the area of the site. GIS programs, including GFW Pro, can easily create these radius circles.

There is an approach

With regards to the CA, our goal is to represent each site individually (the latter 2 options). However due to lack of data on individual sites or to data confidentiality constraints this might not be possible.



(a) Single group centroid point.



(b) A polygon around multiple sites.



(c) A location point for each site.



(d) Polygons for each site.

Figure 1. Different ways to represent the location for a certificate consisting of many sites.

2.3 Permanent and unique ID codes for locations

To keep track of certificates, certificate scheme owners are expected to assign unique and permanent identification (ID) codes to each certificate awarded. These ID codes serve to identify sites if each certificate only has one site. However, where there is more than one site per certificate a coding scheme is needed to uniquely identify each site. For these purposes, farm name or owner name are problematic since they are not always spelled the same way and might exist multiple times.

This is an issue that will need further discussion. There are 2 general approaches of assigning site ID codes:

- Global approach – assign a globally unique code to each site, or at least a code that is unique within the certification scheme.
- Unique within certificate approach – assign a code that uniquely identifies the site within the certificate (but not necessarily unique across all schemes). This approach has the advantage that ID codes can be assigned by the scheme owner however, it has the disadvantage of requiring a globally unique identification of a site to include both the certificate ID and the site ID.

The selection of the best approach depends on at what point in the certificate life cycle the ID codes are assigned. If the ID codes are assigned during data entry into the master database then it is better to let the database generate the ID (first approach). And, alternatively, if they are assigned by the scheme owner then it is more practical for the codes to be unique within the certificate (second approach). It is important that whatever approach adopted is used for all sites in the CA.

Maybe the best approach is for the CA to generate their own site ID codes independently of those used by the scheme owners. There are independent organizations that can generate universal and neutral place ID codes for us that are worth considering. Here are two:

2.3.1 Blunumber

<https://www.blunumber.org/>

“A blunumber is a unique number that identifies people, organizations, places and things.” They are free to register and are accessible to everyone. Bonsucro is one of the Blunumber partners.

2.3.2 GS1 barcodes

Through working with GS1 unique barcodes for sites and/or products can be generated which are readily available for many scanning devices, including smartphones.

<https://www.gs1.org/barcodes>

Both of these approaches will be considered more closely.

2.3.3 Geography-based code

One variation to the global code, is to let the geography define the code. This can be as simple as combining the latitude and longitude into one code. For example, the code “N03808E096494” represents the latitude/longitude coordinates 3.808N and 96.494E, respectively. It meets the criteria of being unique, permanent and can be generated by anyone who knows the location. Since the coordinate value includes 3 digits to the right of the decimal, then it represents a 100m x 100m piece of ground. As long as the sites are at least 100m apart, there will never be 2 sites that have the same code. If more spatial precision is required, then it is easy to add another digit to the code for each coordinate. The disadvantage is a string of 13 characters might be difficult to use.

2.4 Location attribute data

Attribute data are the tabular data that are associated with the site. In the table below we list mandatory attributes, those must be completed for each location, and other optional attributes.

Table 1. List of attributes associates with the certification locations, including the corresponding attribute name used in GFW Pro.

Attribute	Mandatory	GFW Pro attribute ¹	Format	Comments
Cert scheme	Yes	GFW custom code	Text	
Cert ID	Yes	Company code	Text	
Site ID	Yes	Company ID	Text	
Name	Yes	Location name	Text	GFW – must be unique
Latitude	Yes, point only	Latitude	Decimal degrees	?? centroid for polygon
Longitude	Yes. Point only	Longitude	Decimal degrees	?? centroid for polygon
Point/Polygon	Yes	Point/Polygon	Category	
Single/multiple locations	Yes	Custom	Single/Multiple	If records represent one or multiple locations
Crop/Service	Yes	Commodity	Category	
Analysis area	No	Analysis area	Hectares	
Owner name	No	Owner name	Text	
Owner ID	No	Owner ID		
Company name	No	???	Text	
Date certified	No	Company name	Date	N/A if not certified
Certified area	No	Certified area	Hectares	
Admin areas	No	Admin areas	Hectares	Non-production and non-conservation
Production area	No	Production area	Hectares	

¹ GFW Pro is a program developed by WRI for uploading site data using standardized data formats that will be discussed later in these guidelines.

Natural ecosystem set-aside	No	Natural ecosystem set-aside	Hectares	Non-production
Restoration set-aside	No	Restoration set-aside	Hectares	Non-production

3 Collection of location data

Location data can be obtained by either going to the field with a global positioning system (GPS) receiver or visiting the location virtually using an interactive web-based map. Often these approaches are combined; location points and tracks are initially collected using a GPS or smartphone, and then these data are uploaded into a mapping program (such as Google Earth) for viewing, editing and creating the final list or layer of geospatial data, which is stored as a kml/kmz file.

3.1 Field data collection using GPS

The GPS uses satellite signals to calculate and record the current location on the ground. Usually a user assigned point name, a time/date stamp and device ID are also recorded. Now GPS receivers are embedded in many devices, such as cars, smartphones, tablet computers and cameras, moving the use of dedicated handheld GPS devices toward becoming obsolete.

Most GPS enabled devices provide latitude/longitude coordinates with an accuracy of around +/-10 meters. Latitude/longitude points should be collected in decimal degrees² with the number of decimal places representing the given accuracy (e.g. 4 decimal places provides an accuracy of 10 m). Note that in decimal degrees the latitudes in the southern hemisphere and longitudes in the western hemisphere (Americas) must be negative. Therefore, the following latitude/longitude coordinate pair define a location point in Bolivia and provide an accuracy of 10 meters.

-16.9013, -62.0244

GPS coordinates are more accurate when recorded in open areas and away from tall buildings, dense vegetation and steep terrain.

Once GPS data are collected they can be downloaded to your computer in several ways:

- Connect device to computer with USB cable and transfer.
- Send as email attachment.
- Place in a shared cloud space, such as Dropbox.
- Synchronize data between mobile and desktop apps (e.g. Google My Maps).

The common data file formats for GPS data exchange are kml/kmz or gpx files.

² In the configuration of most GPS devices and smartphones, it is possible to set the output of latitude/longitude coordinates in decimal degree format.

3.2 Turn your smartphone into a GPS

There are many mobile apps that essentially turn your smartphone into a GPS enabled mapping tool. The following are just two examples that are free, easy to use, and work offline. Latitude/longitude coordinates can be read directly off the phone or recorded as a waypoint for transferring onto your computer.

3.2.1 GPS Essentials (<http://www.gpsessentials.com>)

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 figure 2). Next, wait for a moment for your phone to pick up satellites and improve accuracy. 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.

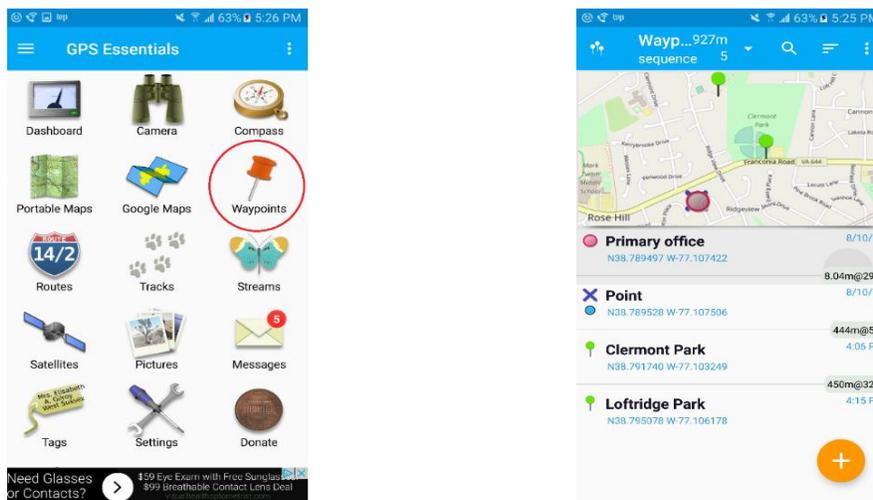


Figure 2. Screen shots of GPS Essentials app.

3.2.2 Google My Maps (<https://www.google.com/maps/about/mymaps/>)

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 standard 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 using the device’s GPS. When you return from the field and connect with the Internet, these mobile maps synchronize with the web-based *My Maps* program, where the data can be exported as a .kml file. The full functionality of the My Maps app does require an internet connection, however, it is possible to collect GPS points while offline.

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 3), 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.

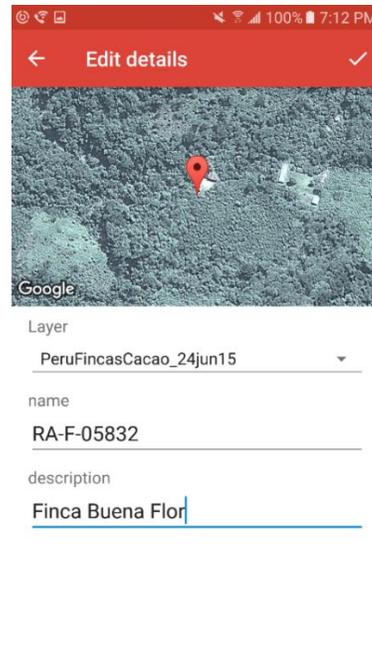
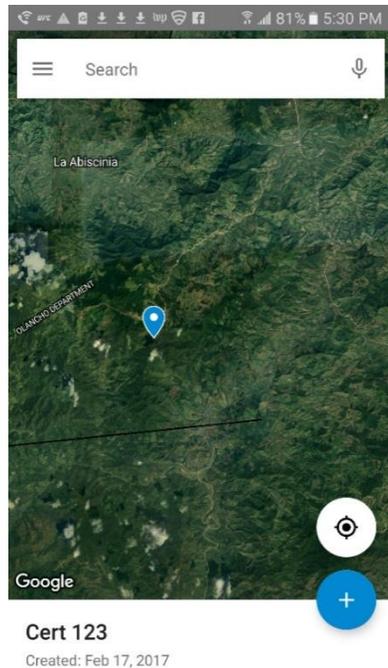


Figure 3. Google My Maps mobile app screen shots.

3.3 Collection of polygon data in the field

Many GPS devices and mobile mapping programs have modules specialized for the collection of polygons in the field. There are 2 ways to do this:

1. **Recording tracks** – this involves activating the tracking (or streaming) feature on the device and then walking the boundary of the unit until you arrive at the place where you started to close the polygon. Save the polygon you just created and insert the ID code somewhere in the polygon name.
2. **Recording vertices or waypoints** – In this approach, you stand on the first corner of the area to be mapped and record a waypoint, inserting the ID into the name of the waypoint field. You then walk to each corner of the area being mapped in sequence and click on the device to register each corner (vertex) of the polygon. Once you return to the starting point you save the polygon or final waypoint, and enter the site ID.

We strongly recommend the second (vertex) approach for the following reasons:

- It does not require walking the exact boundary of the area mapped, as long as a vertex is recorded at each corner.
- Rather than a meandering line, the result is a clean polygon with clear corner vertices.
- The final polygon is simpler and easier to edit because there are less points.

As an alternative to using a specialized polygon collection feature on your device, you can collect points as vertices (approach 1 above) and then use an interactive mapping program to digitize the polygons on top of the points you have collected (see next section).

3.4 Digitizing of location data using interactive maps or GIS

There are numerous publicly available interactive mapping programs (including GIS programs) that display as basemaps with high resolution satellite imagery and allow the user to digitize points and polygons. Here we mention 2 of the most well-known programs: Google Earth and Google My Maps. Both of these programs allow the user to upload existing spatial data (as kml/kmz file), include tools to digitize features and output these features (also as kml/kmz files).

Alternatively, you can use a GIS program; especially if your organization already has one and someone trained to use it. In addition to digitizing, a GIS provides additional options for creating, editing, analyzing and presenting geospatial data in maps. The GIS is especially valuable (probably essential) for managing large datasets and for integrating spatial and tabular (attribute) data.

4 Management and processing of spatial data

Scheme owners, certification bodies and certificate holders should follow these guidelines to start collecting and managing certificate location data, if they are not doing so already. There exists a wide range of computer programs to do this, ranging from simple Excel tables (for points) to GIS.

Here is an overview of the steps that make-up data management process.

1. Define clearly what a “site” represents (e.g. a certificate or single location within a certificate).
2. Develop a scheme for generating unique and permanent ID codes for each site.
3. Define how a site will be spatially represented (e.g. a point or a polygon). If there are polygon data then some type of mapping or GIS program is needed.
4. Define the set of attributes to be collected for each site.
5. Collect site data with GPS, interactive map, or both.
6. Clean and edit site data. Incorporate attributes or manage attributes as a separate table.
7. Upload site data into GFW Pro or GIS for analysis.
8. If you agree to sharing data, export and submit data to ISEAL for inclusion in CA.

4.1 Tabular and point data

The logical choice for the management of tabular and point data is Excel. With the columns populated with latitude and longitude coordinates, these data can be easily mapped using many programs, such as Google Earth, Google My Maps and Bing Maps. To do this you will likely have to first export your data from Excel as a comma separated value (csv) file.

4.2 Geospatial polygon data

Polygon data are more complex than simple geographic points and, therefore, require special programs. These can be the web-based mapping programs mentioned above or more sophisticated geographic information systems (GIS). The feature that separates a GIS from other mapping programs is its ability to manage and analyze associated tabular data. The most common GIS programs are ESRI ArcGIS (<http://www.arcgis.com/>) and the free and Open Source QGIS (<http://www.qgis.org>). Both require significant training and practice to become proficient at using the program.

The common formats for the exchange of spatial data are:

- Common separated value (csv) for tabular data.
- Keyhole markup language (kml/kmz) for point, line and polygon data. This is Google’s default format. Kmz is the same as kml but compressed.
- Shapefile – this is ESRI’s format for exchanging geospatial data, and has become the unofficial standard for the exchange of GIS data. A “Shapefile” is a series of up to 8 files with the same name in which one file has the extension “shp”. It is often zipped into one zip for uploading into mapping programs (such as GFW).
- Gpx is the exchange format used by many GPS devices.

Google Earth (<https://www.google.com/earth/>) is a free mapping program that first must be downloaded and installed. It allows the user to upload and map spatial data in all the standard formats cited above and export as kml/kmz.

Google My Maps (<https://www.google.com/maps/d/>) is a web-based application that can sync with the Google My Maps application on your smartphone. Data can be imported in the following formats: csv, xlsx, kml and gpx, and exported in kml/kmz.

The key ID field is what holds the data in different files together. For example, if tabular data are to be managed in Excel and the spatial features in a mapping package, then it is essential that the ID field be included in both the table and recorded with the spatial feature.

4.3 GFW Pro

Global Forest Watch Pro³ is a free, scalable and easy-to-use management system that any organization can use to manage location data and carry-out standardized analyses by combining with spatial data layers provided by GFW, such as tree cover, tree cover change and fire alerts. GFW Pro users can upload point and polygon data, and securely store these data online for future retrieval and analysis.

With GFW Pro’s secure login system, organizations can have multiple user accounts with one account designated as the “super user” to manage the other user accounts, sharing and other details for their organization. When data are uploaded, by default they are only visible by the user who did the uploading. These data can be shared with their organization if so desired, but not beyond the organization. Options are available for sharing the results of analyses with other GFW Pro users.

GFW Pro is a practical option for ISEAL members and, especially, those that want to or are in the process of collecting location data on certified operations and are not sure how to manage these data. It provides a ready-to-go system for certificate point and polygon data that uses structured data formats and pre-developed land cover change and risk analyses. Data uploaded into GFW Pro can be downloaded in standardized formats for uploading into other systems if so desired⁴.

4.4 Certificate Atlas platform

The goal of the CA is to provide a public facing platform for sharing certificate data in aggregate and, eventually, site levels. What this platform will look like is still under discussion and beyond the scope of these guidelines.

³ GFW Pro has been developed by the World Resources Institute (WRI) and is scheduled to be launched in November/2017.

⁴ Currently only point data can be exported as csv files, however GFW Pro developers are working on tool to download polygon data as ESRI shapefiles or kml.

5 Data disclosure, and security

We all know that data confidentiality, disclosure and security are important and sensitive issues. We also appreciate that data transparency plays an important role in certification. Some producers have security concerns about their location being made public because it might give the competition an advantage or pose a security risk, while consumers want verifiable assurances that the products they purchase are from sustainable sources. When watchdog groups discover destruction of natural habitats, companies want to be able to prove that their supply-chain is not contributing to this destruction. These are just some of the concerns surrounding the issue of data confidentiality and disclosure.

The long-term vision of the CA is to provide this transparency. This is a long process, which begins with the collection and management of location data by the certification schemes and leads to the eventual sharing of these data in a way that strengthens the certification process. In this early stage, the disclosure of certificate is purely voluntary. It is up to each scheme to develop and follow their own data disclosure policies, which are usually formalized in the client agreements between the producer and the certification body. The more certification schemes there are that share their location data with the CA the more useful information it will be available for everyone interested in certification.

Taking into account the sensitivities that some certification standards have around sharing location data, ISEAL's approach is to separate the collection and management of the data with the sharing of data. The procedures and tools presented here, including GFW Pro, are secure. It is the obligation of a credible certification scheme to manage information on what is being certified, and these guidelines should help them do that. Whether or not to share the data with ISEAL through the CA is a separate decision.

5.1 Sharing aggregate data

The sharing of aggregate data on certified activities is a different matter since it does not reveal details on individual certificates or sites. ISEAL proposed that all schemes share the following aggregate statistics with ISEAL and the public:

- Certificate standard or scheme
- Country
- Crop or service
- Number of certificates
- Number of locations or sites
- Total certified areas (hectares)

6 Conclusion and areas of discussion

These guidelines communicate to ISEAL members the technical basis for collecting and managing certificate location data using standard protocols. Through application of these guidelines, ISEAL members can strengthen their capacity to manage certificate location data. WRI's GFW Pro program is presented here as a tool that members can use to upload and manage certificate data.

We encourage those members that are ready to share certificate locations to follow these protocols to submit their locations to ISEAL as either lists (point data), or kml/kmz and shapefiles for polygon data.

As a discussion document, we encourage feedback from ISEAL members. The following are important areas that are expected to receive significant attention and we invite discussion among ISEAL members:

- The relations between the certificate and the site -- how important is it for ISEAL to target collecting the smaller more detailed site data?
- What is the best way to assign permanent and unique ID codes to certificate sites?
- What are the important (mandatory and optional) attributes that should be collected and maintained for each site?
- How should ISEAL address the issues of data confidentiality/disclosure? Should ISEAL attempt to establish standardized guidelines for data disclosure and, if so, what should they look like? What are the concerns and benefits about disclosing certificate site location?
- What aggregate statistics would be reasonable to collect in the CA across all ISEAL certification schemes?
- What are the key features (use cases) that should become part of a public facing Certification Atlas?