FINAL REPORT OF THE CROWDVAL PROJECT

Using Crowdsourcing and Innovative Approaches to Evaluate and Validate ESA's Land Cover Products

LACO-Wiki

Table of Contents

1 Introduction

This is the final report of the CrowdVal project, which provides an overview of the activities that took place in the project.

The objectives of the CrowdVal project were to:

- Develop new innovative sampling schemes that allow a stratification and bias removal via road networks and that take other constraints into account for in-situ data collection;
- Enhance LACO-Wiki and LACO-Wiki Mobile with the new sampling strategies, functionality for opportunistic map evaluation on the ground, and the addition of auxiliary data sets including Flickr geo-tagged pictures and time series of NDVI;
- Create an open source version of LACO-Wiki Mobile;
- Demonstrate the enhanced tools through crowdsourcing data collection campaigns (online and in-situ) to validate the first land cover map of Africa at a 20m spatial resolution; and
- Investigate the possibility of developing a business model around an open source version of LACO-Wiki Mobile with a payment model around access to enhanced features, e.g. additional data sources, commercial satellite imagery, increased sample size, etc.

To achieve these objectives, the initial project meeting of the consortium was focused on discussing the user requirements for LACO-Wiki and LACO-Wiki Mobile. These were provided to ESA in August 2018 as **Deliverable D1: Report with technical specifications for new features in LACO-Wiki and LACO-Wiki Mobile** (see Annex A for a list of the project deliverables). The sections that follow describe how each of these objectives have been achieved and any deviations from these objectives, with appropriate justification.

2 Innovative Sampling Schemes

The first objective of the CrowdVal project was to add innovative sampling schemes to LACO-Wiki, which consisted of two parts: a) add an algorithm for sampling along a road network (to be implemented by Gisat); and b) add an algorithm for consideration of additional accessibility constraints (to be added by CS). As outlined in the Mid-Term Report (MTR), it was decided to only implement the first option as it was the most practical from a user perspective. This is based on the experience from ground sampling conducted by IIASA during the first campaign in Kenya, which clearly showed difficulties and constraints in collecting in-situ samples at large distances from roads and cart tracks. The other constraints were considered to be less important and the focus was directed to providing samples on a road network. Hence, we diverted the resources of CS from this task (originally Work Package 3) to mobile app development (Work Package 4), since CS staff members have specific expertise in this domain.

To achieve this objective, Gisat built a service for sampling along a road network, which is also **SW1: Code for new sampling strategy that takes road networks into account** (see Annex A for a list of the project deliverables). This service was then integrated into LACO-Wiki, where two new options were added: (i) random point sample on the road network; and (ii) random pixel sample on the road network. These options appear when a raster map is uploaded to LACO-Wiki (or when using an existing shared raster map already in LACO-Wiki) as shown in Figure 1. The algorithm sends the bounding box of the raster map to the service developed by Gisat, which uses OpenStreetMap (OSM) as the road network and the Overpass API to extract the roads from OSM. Note that the Overpass API[1] can be unstable so LACO-Wiki calls the API at least 10 times if it fails during the creation of the sample. Moreover, we have set a maximum limit of 10K km^2 as the size of the bounding box as larger sizes cause problems due to insufficient memory. However, realistically, data collection on a road network would be in a relatively small area, e.g., a district, so such a limit is also practical.

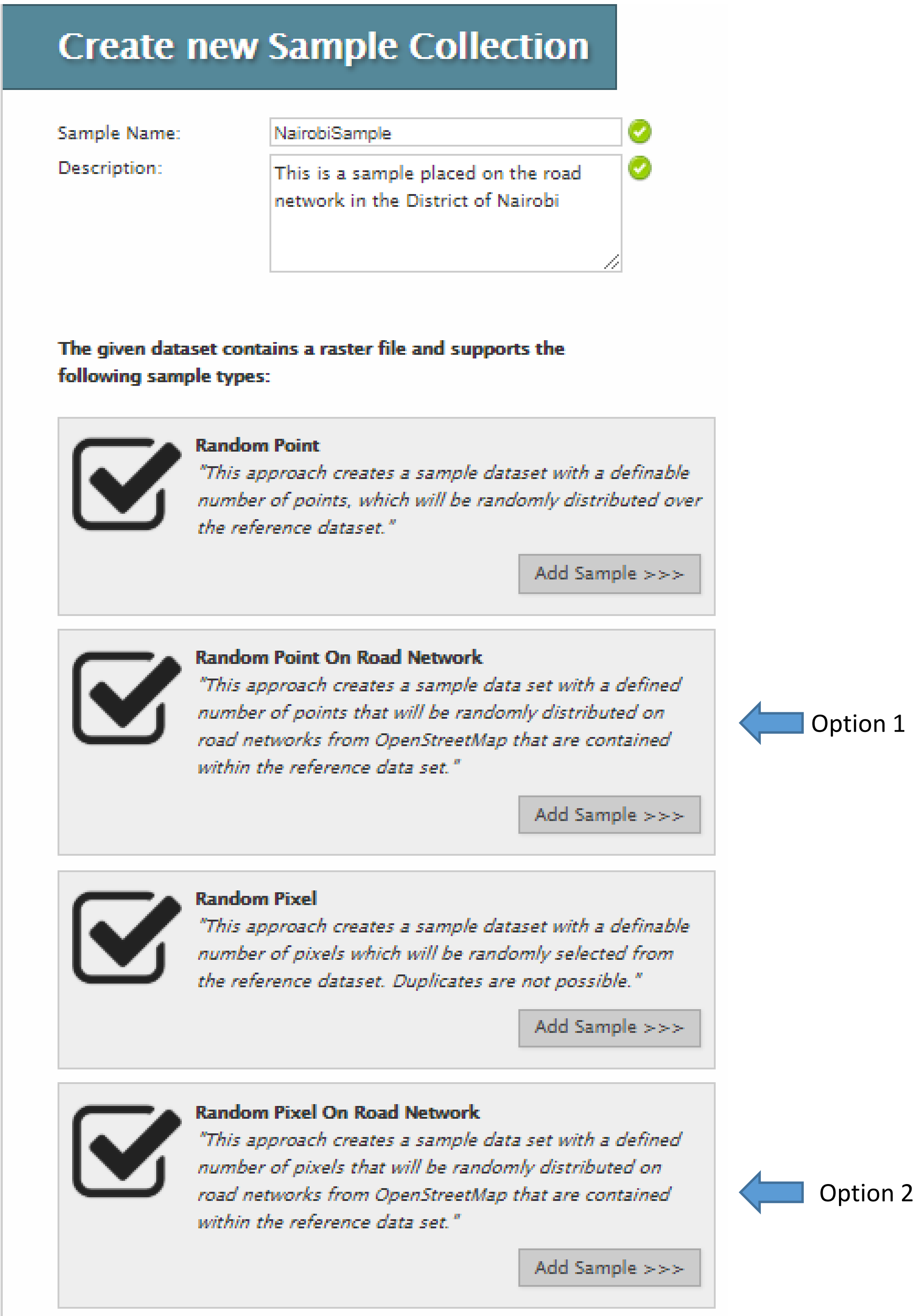
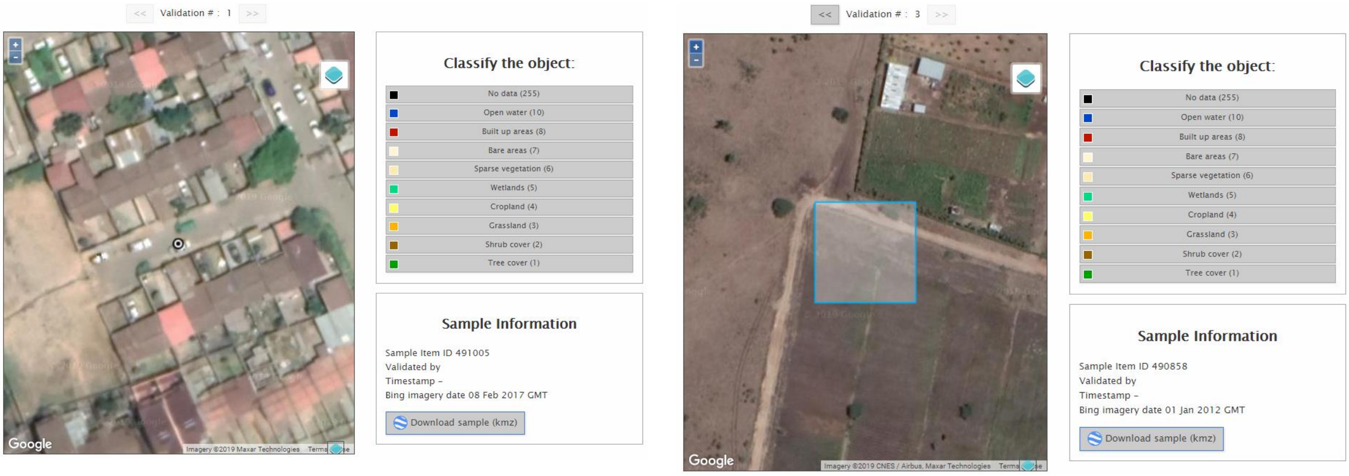


Figure 1: Two new options in LACO-Wiki to produce a random point and random pixel sample along a road network

[1] https://wiki.openstreetmap.org/wiki/Overpass_API

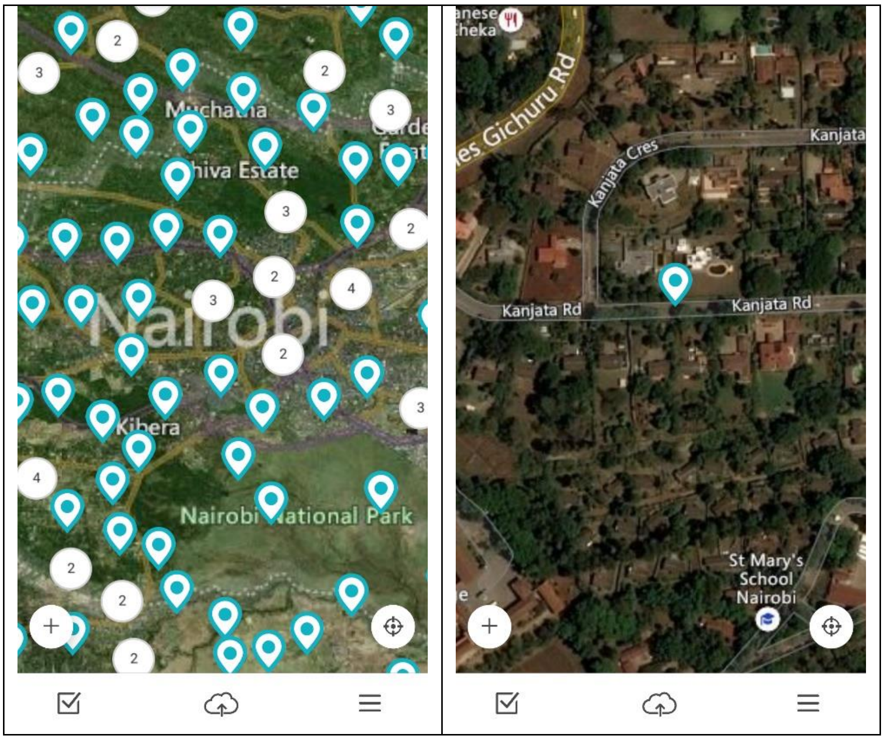
Figure 2 shows two examples of the road network sampling algorithm applied to random points and random pixels along a road network in visual interpretation sessions of LACO-Wiki. Clearly the algorithm makes more sense for in situ data collection, but this simply demonstrates that the algorithm works.



(a) (b)

Figure 2: An example of (a) a random point and (b) a random pixel located along a road network in LACO-Wiki

Hence the algorithm would more likely be used for in situ data collection with LACO-Wiki Mobile. Figure 3a shows an example of points located along roads in Nairobi in LACO-Wiki Mobile, which also demonstrates this feature working in the mobile app. Figure 3b shows an example when zoomed into the map.



(a) (b)

Figure 3: An example of random points located along a road network in LACO-Wiki Mobile for (a) a large area in the Nairobi district and (b) zoomed in to one location to see one point more clearly

It is possible to use data collected along a road network for statistically rigorous land cover validation, but this would involve determining the bias and then correcting for it. Instead it makes more sense to collect the data along a road as training data for classification algorithms or for verification of classes in a land cover map.

3 Enhancements to LACO-Wiki and LACO-Wiki Mobile

The second objective of the CrowdVal project was to enhance LACO-Wiki and LACO-Wiki Mobile with new functionality, which is described in more detail in the sections that follow.

3.1 Enhanced plausibility validation (LACO-Wiki Mobile)

In LACO-Wiki, there are three ways to create a validation session:

- **Blind validation:** The user is shown the location of the pixel (point or polygon) on the satellite image and is given a set of land cover classes to choose from. However, the user has no information about the land cover type from the land cover map.
- **Plausibility:** Same as blind except that the land cover class from the map is shown to the user. The user must then confirm the land cover class from the map or indicate that it is incorrect (based on visual interpretation).
- **Enhanced plausibility:** Same as plausibility but the enhanced version refers to the ability to choose an alternative land cover class if the land cover map is incorrect.

All three types of validation session were present in the online version of LACO-Wiki prior to the start of the CrowdVal project but LACO-Wiki Mobile was only capable of handling blind and plausibility validation sessions. During the CrowdVal project, enhanced plausibility validation was added to LACO-Wiki Mobile. Figure 4 shows an example of screen dumps from the LACO-Wiki Mobile showing this option.

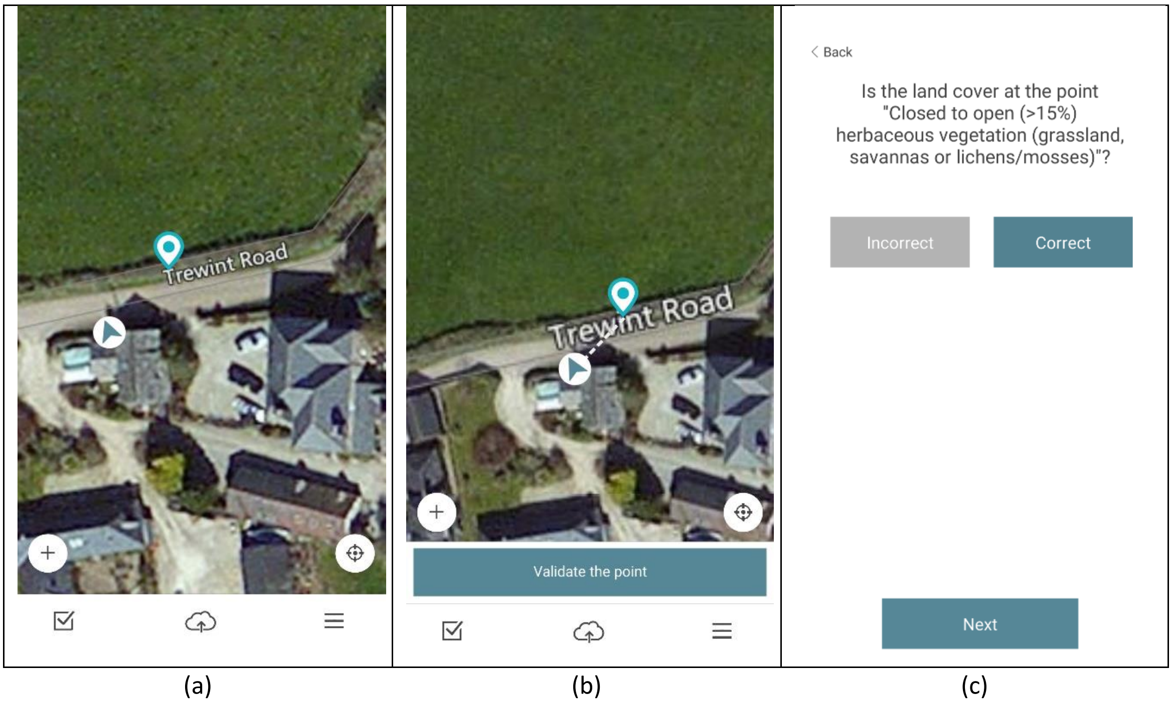


Figure 4: Example of enhanced plausibility in LACO-Wiki mobile showing screenshots of (a) a point for validation shown on the map; (b) ability to validate the point once the user is close enough; and (c) validation of the land cover at that location.

Figure 4a shows a point on the map in the sample. If the user is close enough to the point, clicking on the point will allow the user to validate the point (Figure 4b). The user is then told the land cover at that location. In this instance the land cover is correct. If the land cover had been incorrect, then the user would be given a list of all land cover types to choose from in order to correct the class.

3.2 View very high-resolution imagery in offline mode (LACO-Wiki Mobile)

This feature was not originally in the proposal, but it was deemed to be an important feature when using the LACO-Wiki Mobile in situ, particularly in countries where the cost of mobile data is high and internet connections are not available in the field. This feature was implemented by CS instead of sampling strategies that take additional constraints into account (see section 2). Figure 5 shows how this is implemented in the LACO-Wiki Mobile app. The option to download an offline map is in the settings (Figure 5a). The map can then be viewed when offline (Figure 5c), allowing users to validate the sample points in the field when no data or internet connection is available.

Figure 5: Accessing very high-resolution imagery from LACO-Wiki Mobile. Screenshots from LACO-Wiki Mobile showing (a) the Download Offline Map option; (b) choosing the area for downloading; and (c) viewing the maps while offline.

3.3 Multiple validations at the same location (LACO-Wiki and LACO-Wiki Mobile)

For reasons of quality assurance and the application of approaches such as majority voting to the validation data set, having multiple validations at the same location is a useful feature. The way to do this involves the following steps:

- Create an initial sample of any type, i.e., random, stratified, etc.;
- Download the created sample as a shapefile from LACO-Wiki;
- Create a second sample using the augmentation options that have been added to the online version of LACO-Wiki (see Figure 6a) as part of the CrowdVal project. Load the shapefile (see Figure 6b) downloaded in the previous step; and
- Repeat the previous step to create the number of sample sets that correspond to the number of times you want a location to be validated.

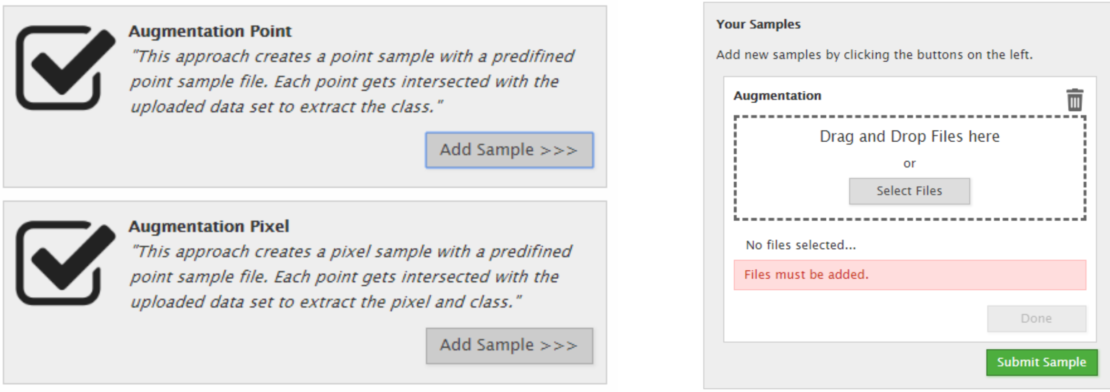


Figure 6: Implementing multiple validations in LACO-Wiki using (a) the new augmentation feature to (b) upload a sample that was created initially and downloaded using LACO-Wiki

The process outlined above produces multiple samples of the same data set. These are then turned into multiple validation sessions which are assigned to (or shared with) different individuals, who then carry out the validation. This can apply equally to multiple validation sessions for the web-based version (visual interpretation) or data collection in situ using LACO-Wiki Mobile.

3.4 Opportunistic data collection (LACO-Wiki Mobile)

Opportunistic data collection is a new, key feature of LACO-Wiki Mobile. In addition to collecting data at the pre-specified sample locations (created as part of a sample in the online version of LACO-Wiki), users can use LACO-Wiki Mobile to collect land cover data at any location, i.e., opportunistically.

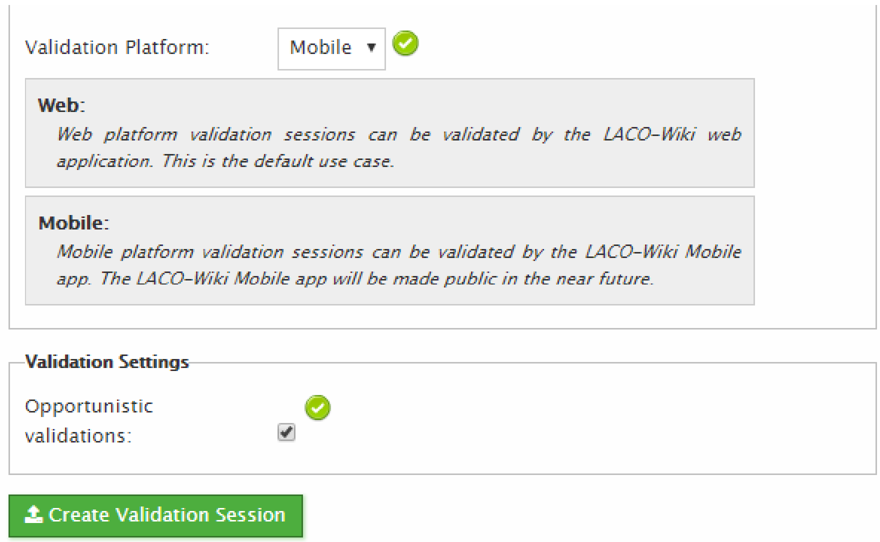


Figure 7: Enabling opportunistic validations during a validation session for in-situ data collection via LACO-Wiki Mobile

The only limitation is that the opportunistic samples must be within the bounding box of the sample created in the online version of LACO-Wiki. Figure 7 shows an example of activating the opportunistic sample data collection feature in LACO-Wiki Mobile. Figure 8 shows how opportunistic data collection works. The + sign at the bottom of the screen (Figure 8a) activates opportunistic data collection. The user is then taken through a series of steps to record the location, choose the land cover type and take photographs of the land cover.

Figure 8: Opportunistic, in situ land cover data collection using LACO-Wiki Mobile. (a) Screen showing sample points (blue markers) and the + button on the bottom left; (b) After pressing +, the user is asked if they want to validate at their location or at a location to be selected on the map; (c) If select on the map is chosen, the user is asked to place a marker on the map; (d) Once the marker is placed, the user can add a land cover observation by pressing Validate location; (e) the user chooses the land cover type; (f) and then takes pictures in 4 cardinal directions; (g) a final screen is displayed informing the user that the point has been added.

3.5 Addition of NDVI profiles (LACO-Wiki)

The display of NDVI profiles is another new, key feature that was added to the online version of LACO-Wiki, which can greatly aid visual interpretation, particularly for those land cover classes that are difficult to differentiate from the satellite imagery alone or which have a seasonal component.

The NDVI profile must be enabled when creating a validation session. An example is shown in Figure 9 in which a check box is turned on if NDVI profiles are to be displayed. After the session is created and the validation session entered, two sets of NDVI profiles will appear below the main validation area, as shown in Figure 10. The chart on the left shows a time series of 32-day composites from Landsat 7 and 8, a 16-day composite from MODIS and daily time series from the PROBA-V satellite. On the right is the average NDVI from each of these sources.

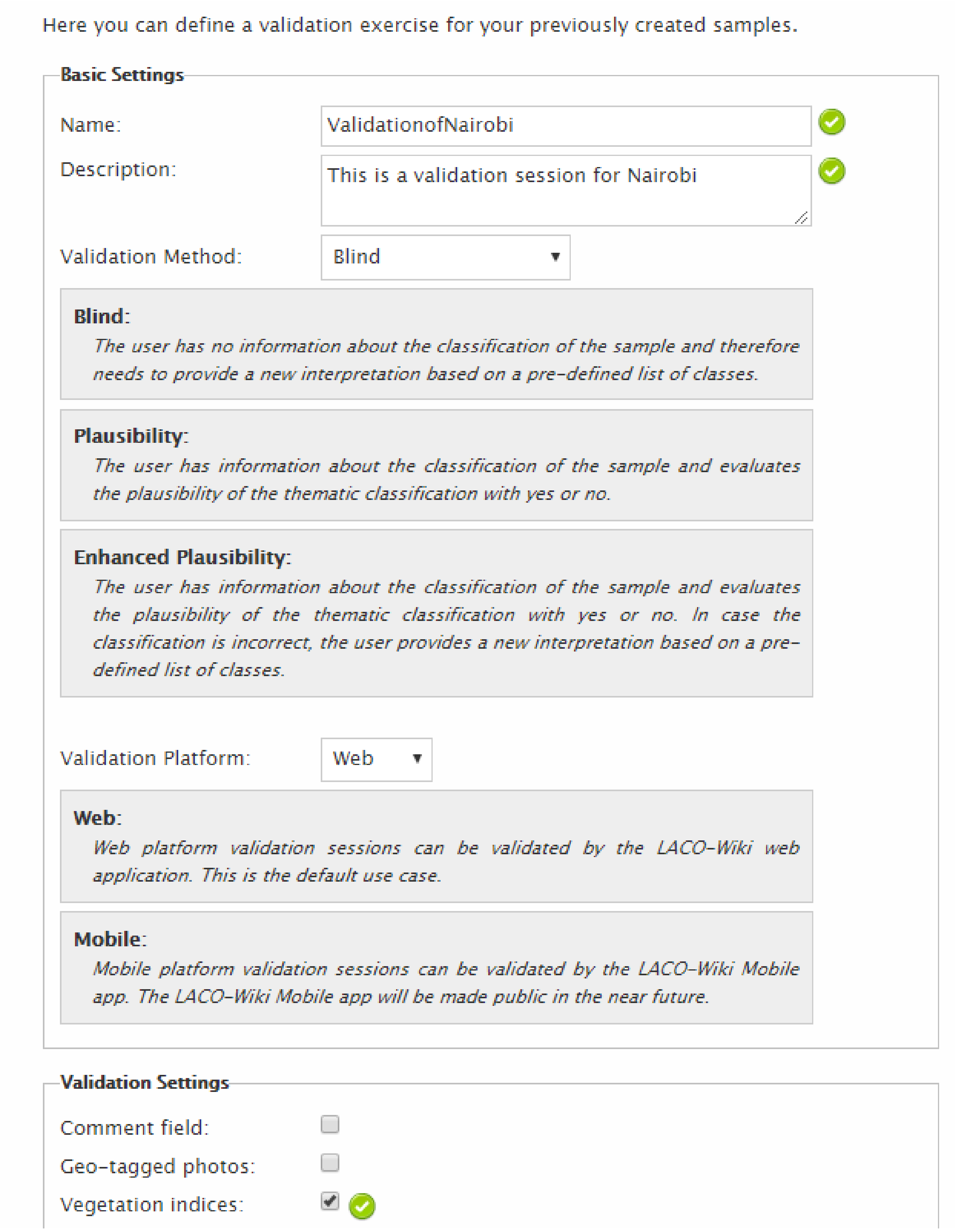


Figure 9: Enabling the NDVI profile feature during the creation of a LACO-Wiki validation session

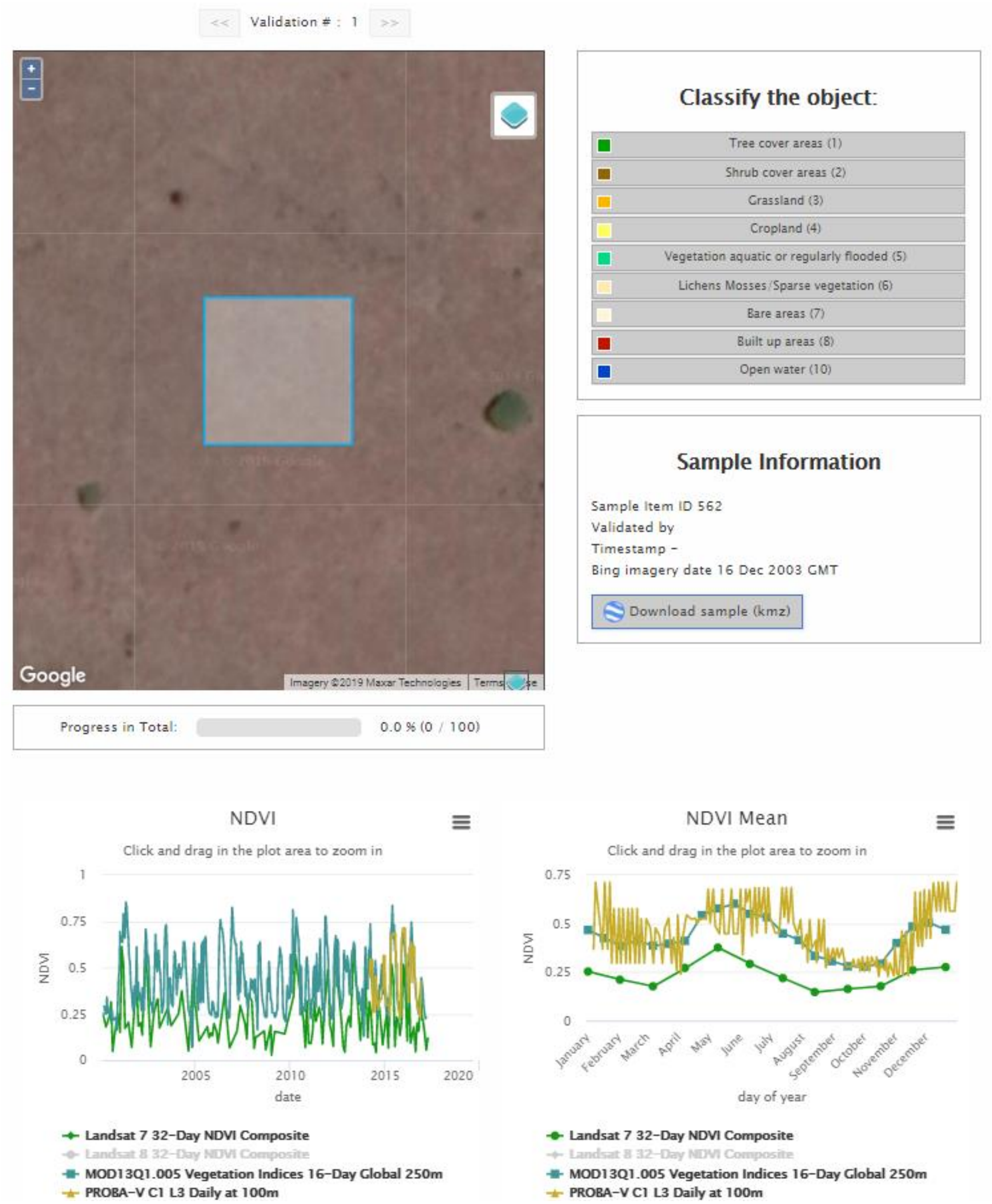


Figure 10: Screenshot of the NDVI profiles that are displayed during a land cover validation session in LACO-Wiki

Note that in the proposal, NDVI profiles would also be added to the mobile app. However, it was decided that this feature does not make sense for the mobile app for two means reasons. The first is practical because of the size of the mobile phone screen, which is too small for meaningful display of NDVI graphs. The second is related to the need for such an option. When doing visual interpretation using satellite imagery, NDVI profiles are very useful for differentiating between certain land cover types. For in situ data collection, the use of NDVI is not relevant because the land cover type is clear from the ground.

3.6 Addition of geo-tagged photographs (LACO-Wiki)

To aid in visual interpretation, geo-tagged photographs from Flickr have been added to LACO-Wiki. Similar to the NDVI profiles, this feature must be turned on during the creation of a validation session as shown in Figure 11.

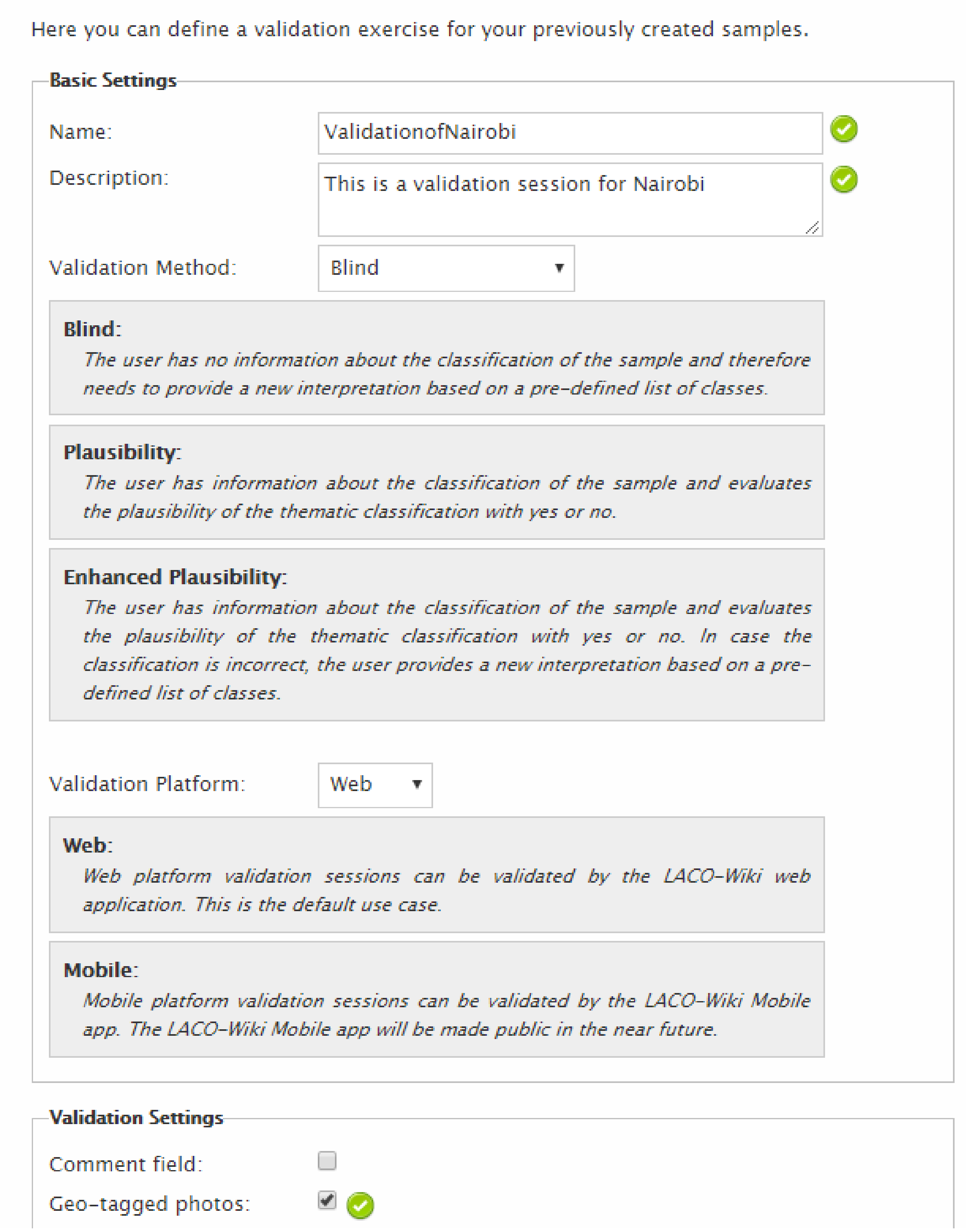


Figure 11: Enabling the geo-tagged photo feature during the creation of a LACO-Wiki validation session

If a geo-tagged photograph is present, it will appear as a yellow dot on the satellite imagery. In most cases, users will need to zoom out to see whether photographs are present. Figure 12 shows an example of zooming out, which shows the pixel to be validated on the left near the top of the screen (small blue box) and a yellow dot for the presense of a geo-tagged photograph. Clicking on one of the dots displays the photograph, which shows a Grassland type of land cover.

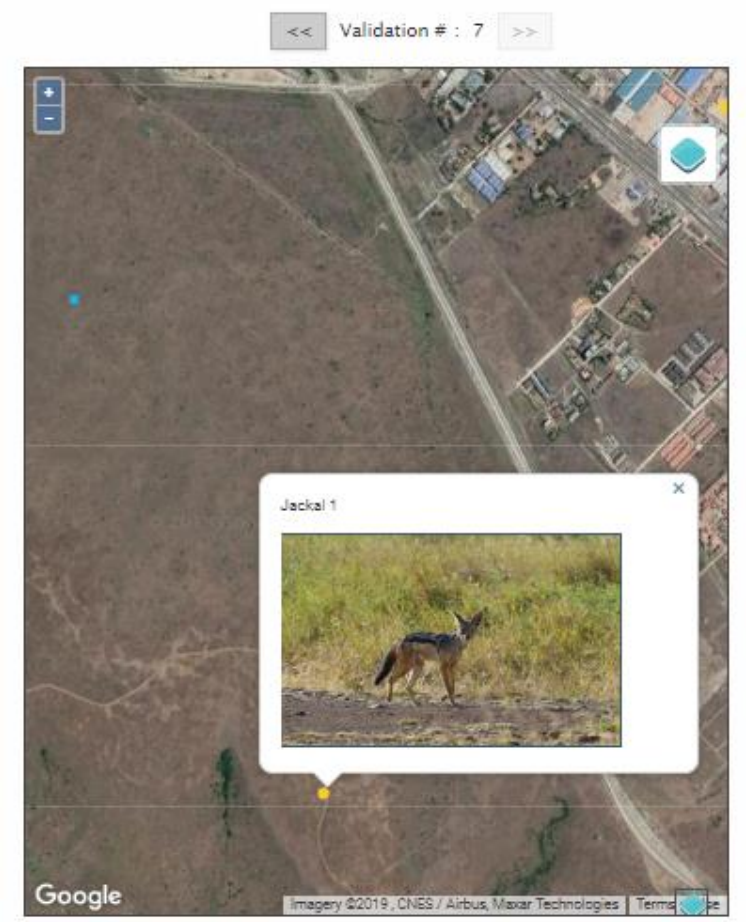


Figure 12: A geo-tagged photograph in the vicinity of the pixel being validated

Note that in the proposal, geo-tagged photographs would also be added to the mobile app. However, it was decided that this feature does not make sense for the mobile app as the user is already on the ground and can see the landscape. Instead it would make more sense to collect geo-tagged photographs with LACO-Wiki Mobile, which can eventually be added to LACO-Wiki as an additional source of information for visual interpretation of satellite imagery.

3.7 Addition of Terms of Use/Privacy Policy (LACO-Wiki)

The Terms of Use of LACO-Wiki were originally listed on the About page of LACO-Wiki. During the MTR in February 2019, ESA flagged the need to have Terms of Use for LACO-Wiki that users must agree to the first time they use the application. Both Terms of Use and the Privacy Policy (compliant with the European Union's General Data Protection Regulation) were added to LACO-Wiki. Figure 13 shows the Terms of Use displayed to the user the first time that they log in. At the bottom of the Terms of Use is a link to the Privacy Policy; hence, users agree to both when they log in the first time. The user can view the Terms of Use and Privacy Policy at any time from the About page of LACO-Wiki:

https://laco-wiki.net/en/About

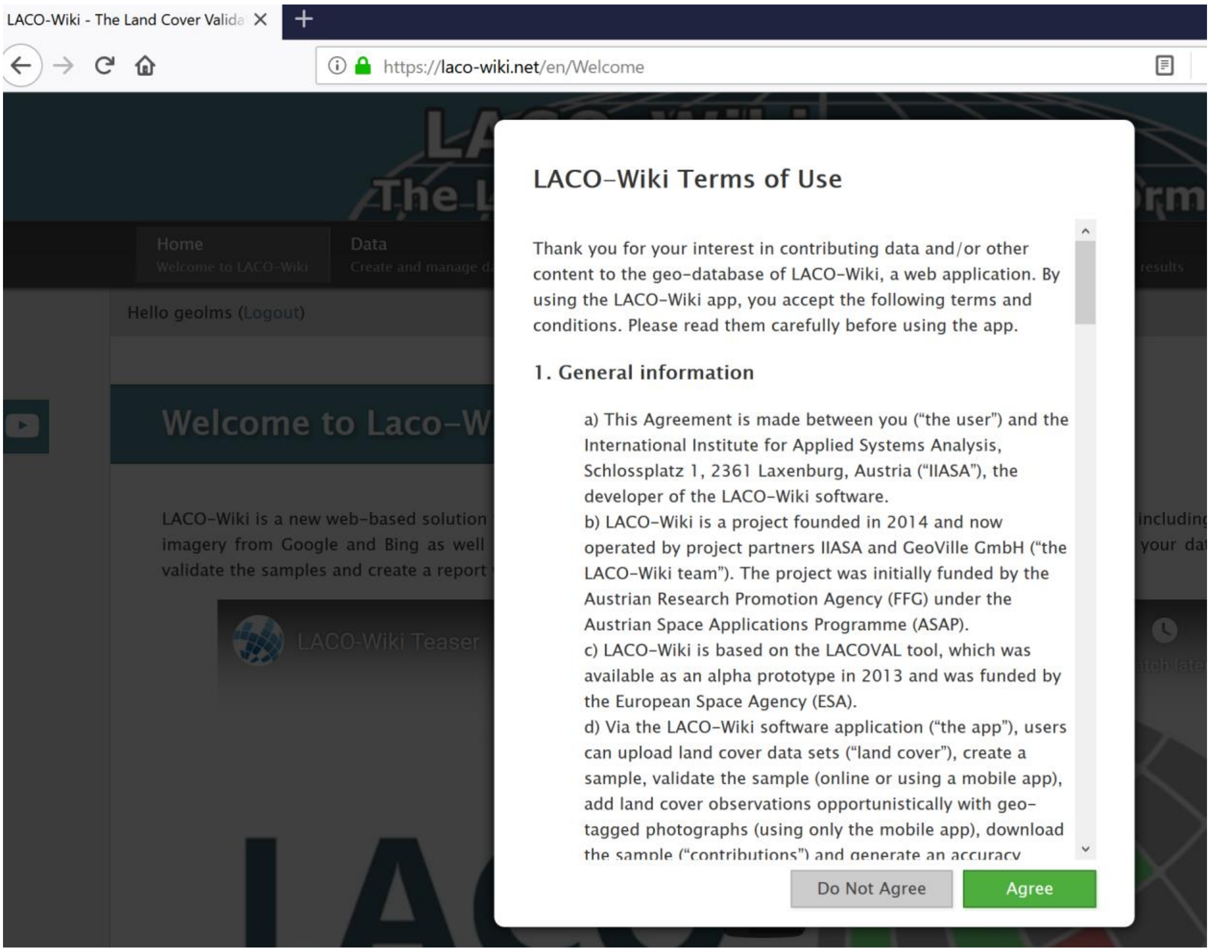


Figure 13: Terms of Use shown in LACO-Wiki the first time that a user logs into the system

3.8 Disabled guest access (LACO-Wiki)

During the MTR meeting in Frascati (28 Feb 2019), ESA noted that one could log into Geo-Wiki using guest access and then automatically be logged into LACO-Wiki. We have since disabled guest access to LACO-Wiki, which now requires logging in with an authorized Geo-Wiki, facebook or Google account.

4 Open Source Version of LACO-Wiki Mobile

The code for LACO-Wiki Mobile is openly available in the github repository at this location (please copy URL into a browser):

https://github.com/iiasa/LACO-Wiki_Mobile_Unity

released under a GNU General Public License version 3.0. This license means others can copy, distribute and/or modify the software. Note that the delivery of this code as open source is also deliverable SW4 of the CrowdVal project (see Annex A).

5 Crowdsourcing Campaigns/Validation of ESA African Land Cover Map

Another objective of the CrowdVal project was to demonstrate the use of LACO-Wiki and LACO-Wiki Mobile in crowdsourcing campaigns in African countries. The data collected during these campaigns were then used to validate the ESA 20m prototype land cover map for these countries. The achievement of these objectives is described below.

5.1 Crowdsourcing campaigns

Details of the crowdsourcing campaigns undertaken as part of the CrowdVal project are provided in **D3: Report on the crowdsourcing campaigns** (as listed in Annex A). Table 1 provides an overview of the campaigns that were run during the project along with the number of points collected in situ and online with LACO-Wiki.

Table 1: Overview of the campaigns and data collected

Campaigns	Date of workshop	Date over which data were collected	Number of points collected in situ	Number of points collected online
Kenya	27-28 Sep 2018	Until Mar 2019	766	10636
Gabon	12-13 Dec 2019	Until Apr 2019	293	1949
Ivory Coast	5-7 Feb 2019	Until Apr 2019	686	6128
South Africa	N/A	June 2019	N/A	92000

In the case of South Africa, the data collection was done internally by IIASA experts. The in situ data and the data collected online are available in the CrowdVal branch of Geo-Wiki (https://www.geo-wiki.org), where they can be visualized or downloaded. This is deliverable Data1 (see Annex A).

The three campaigns in Kenya, Gabon and Ivory Coast provided some interesting lessons on visual interpretation and data collection in the field. In Kenya, there were very few participants who had local knowledge of land cover. Hence training in visual interpretation by an expert (Dr Myroslava Lesiv) was critical in improving the ability of participants to recognize landscapes. In contrast, in Ivory Coast, local knowledge proved to be useful in classifying certain landscapes. Hence good training is essential for the development of high-quality training and validation data sets. However, local knowledge can also be useful for the identification of some land cover types in a local context. This has implications for the classification algorithms applied to satellite imagery, as some features may not be picked up or would simply be misclassified if there was insufficient training or local knowledge.

The mobile app, although a useful tool for data collection, had some limitations. The first was technological as not all participants had smart phones capable of running the app. Difficult landscapes also made it hard to reach some of the points. Hence there is a tradeoff between in situ data collection based on a rigorous statistical sample (and the inability to reach some of these locations) and more opportunistic in situ data collection, where the latter could be used for training algorithms rather than validating a land cover map. Hence the mobile app has more applicability for the collection of training data. The algorithm to place sample points along the road will also facilitate in situ data collection in the future.

5.2 Validation of ESA African land cover map

The validation of the ESA 20 m African land cover map is described in detail in **D4: Accuracy assessment of 20 m land cover map of Africa** (see Annex A for the complete list of deliverables). The data collection campaigns described in D3 were the source of the validation data, where the validation was based on a systematic sample collected using the LACO-Wiki online validation tool. Where possible, locations were validated at least twice, and any disagreements were resolved by one or more experts in order to produce high quality validation data sets for Kenya, Gabon, Ivory Coast and South Africa.

The overall accuracies for Kenya, Gabon, Ivory Coast and South Africa are 56%, 91%, 47% and 44%, respectively. The values for Kenya and South Africa are primarily due to the misclassification of large areas of grassland and shrub cover. There is also generally overestimation of cropland. The value for Ivory Coast is because the land cover is highly fragmented, and therefore, it is a very difficult country to map using remote sensing. Gabon, on the other hand, had very high accuracy, due primarily to the fact that the country is covered largely by forest cover, which is an easy class to map with remote sensing.

In addition to overall accuracy and producer's and user's accuracy, maps of spatial accuracy were provided in the deliverable. These maps were made possible because of the dense validation sample collected as part of the CrowdVal project. These maps can used in future land cover mapping projects to understand geographically where land cover has been mapped poorly in the past and which classes this pertains to. This could also aid the collection of training data, directing more efforts to those areas that were mapped poorly in order to help improve the results produced by the classification algorithms. Other suggestions for improvement are provided in deliverable D4, e.g., using LACO-Wiki and LACO-Wiki Mobile to increase the size of the training data set rather than relying on training data from existing coarse resolution land cover maps as well as the use of additional data sets, e.g., IIASA's field size map, maps of human impact, geo-tagged photographs, etc.

6 Business Models and Future Developments

The code for LACO-Wiki Mobile is currently open and can be used and/or modified by any third party under the GNU Public License v3. This license only concerns free access to the code and does not preclude any third party from charging money for using the app or building a business model around the app.

Some potential business models and future developments include the following:

- Project partner CS has developed algorithms for more accurate GPS location in mobile applications. They could add these algorithms to the app and charge users a fee per point collected (or a subscription fee) if they want more accurate GPS coordinates. This would make the app more relevant in those situations where very accurate locations for the data collection are required.
- Very high-resolution satellite images (e.g., Pleiades), commercial orthophotos or data from Planet could be added to the app via LACO-Wiki; users would only be able to access this imagery if they paid a subscription fee or a cost per square km.
- The sample size for the number of points allocated on a road network is currently limited to 1000 and an area of 10K km^2. These limits are partly due to the computational resources on which the LACO-Wiki server sits. A business model could be built around increasing these limits on a subscription fee basis, which could finance an upgrade to the hardware infrastructure on which LACO-Wiki sits.
- Other sources of data could be added to LACO-Wiki Mobile, e.g., geo-tagged photographs from commercial or other sources, addition of other background layers, etc., which could be financed through sub-contracting the work via IIASA.
- LUCAS data collection through Copernicus costs around 36 Euros per point. A version of LACO-Wiki Mobile could be adapted by a commercial company to collect land cover that could either complement LUCAS (providing a denser sample) or be collected more regularly (every year rather than every 3 years) if there was demand from EU member states, at a price that would be significantly less than the LUCAS exercise.

- At present, the road network sampling algorithm works for the generation of random samples along a road. Stratified sampling could be added, which is possible using the Gisat service but is not currently implemented in LACO-Wiki. This could be financed by through sub-contracting the work via IIASA.
- Animal monitoring for conservation purposes could benefit from our application. Animal presence could be captured and provided to entities such as environmental associations or national parks. Such entities could support the costs associated with fine-tuning of the app and the server maintenance.
- LACO-Wiki could help to revitalize the GOFC-GOLD community where FAO could play a role in this development. The final CrowdVal meeting at FAO could provide the basis for discussing these opportunities.
- Further developments could also include repeating the field-based exercises as researchers at the Université Félix-Houphouët-Boigny in Abidjan expressed interest in future workshops for the purpose of training students in visual interpretation and in situ data collection. This win-win approach could serve as a basis to maintain networks of local experts that could be activated in future land cover mapping and validation exercises.
- Similarly, we could promote the materials more generally for adoption in university curricula related to remote sensing and land cover classification/validation.

These business models and potential developments will be considered as part of following up on future opportunities for LACO-Wiki and LACO-Wiki Mobile.

7 Conclusions

This report has outlined the activities that took place in the CrowdVal project, in particular, how the objectives of the project were achieved. Overall, the project was intended to make technological advances in the LACO-Wiki online validation tool and in LACO-Wiki Mobile, where the latter is now open source and can be used by any third party. These tools were demonstrated in some African countries, and the data collected through LACO-Wiki were then used to validate the ESA 20 m African land cover map. The LACO-Wiki tools represent powerful ways to not only validate land cover maps but also to collect training data for future land cover map developments. These can be exploited in a number of ways as outlined in the business models and future developments section of this report.

Annex A: Deliverables of the CrowdVal Project

Below is the list of deliverables associated with the CrowdVal project.

Deliverable	Name	Description
D1	Report with technical specifications for new features in LACO-Wiki and LACO-Wiki Mobile	Output from WP2. Delivered to ESA August 2018.
D2	Training manual for data collection (English and French)	Output from WP4. Delivered to ESA in the MTR (Feb 2019).
D3	Report on the crowdsourcing campaigns	Output from WP5. Delivered to ESA in August 2019.
D4	Accuracy assessment of 20 m land cover map of Africa	Output from WP5. Delivered to ESA end of August 2019.
SW1	Code for new sampling strategy that takes road networks into account	Source code available in github at: https://github.com/gisat/crowdval_sampling_tool. Please copy the URL to a browser. If there are problems accessing the github site or for permission to gain access, please contact, contact Igor Kratochvíl at igor.kratochvil@gisat.cz. Delivered to ESA end of August 2019.
SW2	New version of LACO-Wiki	Online version available at: https://www.laco-wiki.net
SW3	New version of LACO-Wiki Mobile	Mobile version available from: https://play.google.com/store/apps/details?id=com.iiasa.lacowikimobile https://apps.apple.com/us/app/laco-wiki-mobile/id1465081846?ls=1
SW4	Open source version of LACO-Wiki mobile	Source code available in github at: https://github.com/iiasa/LACO-Wiki_Mobile_Unity
Data1	Training and validation data collected from 1 online and 3 in situ crowdsourcing campaigns	Available for viewing and downloading from the CrowdVal branch of Geo-Wiki (https://www.geo-wiki.org)