COLLECT EARTH ONLINE

Institution & Project Creation Manual

The CEO Team May 20, 2020


USAID
FROM THE AMERICAN PEOPLE

NASA

SERVIR

Food and Agriculture
Organization of the
United Nations

openforis

SilvaCarbon


Spatial Informatics Group

Google

GTAC

COLLECT EARTH
ONLINE

Table of Contents

Welcome to Collect Earth Online

Collect Earth Online, or CEO, is a free and open-source image viewing and interpretation tool, suitable for projects requiring information about land cover and/or land use. CEO enables simultaneous visual interpretations of satellite imagery, providing global coverage from MapBox and Bing Maps, a variety of satellite data sources from Google Earth Engine, and the ability to connect to your own Web Map Service (WMS) or Web Map Tile Service (WMTS). The full functionality is implemented online, no desktop installation is necessary.

CEO allows institutions to create projects and leverage their teams to collect spatial data using remote sensing imagery. Use cases include historical and near-real-time interpretation of satellite imagery, data collection for land cover/land use model validation.

Collect Earth Online (CEO) is available at https://collect.earth/.

This manual provides information for institution and project administrators, including how to set up an institution, how to set up a project, and project management suggestions. Part 1 provides a quick overview of the Collect Earth Online platform, so administrators can set up their own user accounts. Part 2 details how to set up and manage an institution in CEO. Part 3 discusses how to manage custom imagery feeds; if your institution will only be using the standard imagery from MapBox and Bing Maps, you can skip this section. Part 4 discusses how to create a data collection project, Part 5 discusses project review, and Part 6 introduces the Geo-Dash, which provides additional information for uses during data collection. Part 7 explains how to manage a published project. If you are interested in setting up a CEO project for validation (e.g. of a land cover model), Part 8 discusses project creation for these specialized projects.

This manual uses a few formatting standards for ease of use. Clickable links are shown **[like this]**. The names of pages, like the **Home** page, are bolded.

The icon is used to indicate use cases. The icon is used to indicate new, unique, and or useful functionality in CEO. Finally, CEO is a live tool that is under continuous development. Workarounds for current issues are shown in boxes like this:

> This is a description of a current challenge and how to work around it.

And planned features are shown in boxes like this:

> This is a description of a planned feature.

Part 1: Quick Introduction to Collect Earth Online

A. Setting up your account

1. In your browser window, navigate to https://collect.earth/. CEO supports Google Chrome, Mozilla Firefox, and Microsoft Edge.
2. Click **[Login/Register]** on the upper right.
3. To set up a new account, click on **[Register a new account]** and follow the instructions.
4. When you have an account, login with your email and password.
5. If you forget your password, click on **[Forgot your password?]** and follow the instructions.

B. Website features

1. On a desktop browser, you can access the **Home**, **About**, **Support**, and **Account** pages from the top menu bar.
   i. The **Home** page includes information about institutions, published projects, and a map showing locations of existing projects.
   ii. The **About** page summarizes information about CEO.
   iii. The **Support** features Collect Earth Online Manuals, Tutorials, and a Collect Earth Online Demo. This page also includes links for bug reporting and forums to ask for help.
   iv. The **Account** page lists information such as user statistics and allows users to update their account settings.

> The data collection portion of the website does not work on mobile (question/navigation sidebar is not accessible). On a narrow computer/laptop/tablet screen, or if your browser window is not over 958 pixels wide, the data collection screen is located beneath the map and you will need to scroll down in the browser window to find it.

> CEO will be adding dynamic "My Institutions/My Projects" tabs along with more detailed Account pages. This will make it easier to access your institutions and projects.

C. Report an issue and request new features

When you click on **[Support]** at the top of the webpage, there is a link to the GitHub issues page. This page is also available at: https://github.com/openforis/collect-earth-online/issues.

If you discover any of the Collect Earth Online functions are not working properly or would like to suggest an additional feature, you can use this page to log an issue or suggestion. Once logged, these messages go directly to the Collect Earth Online developer team.

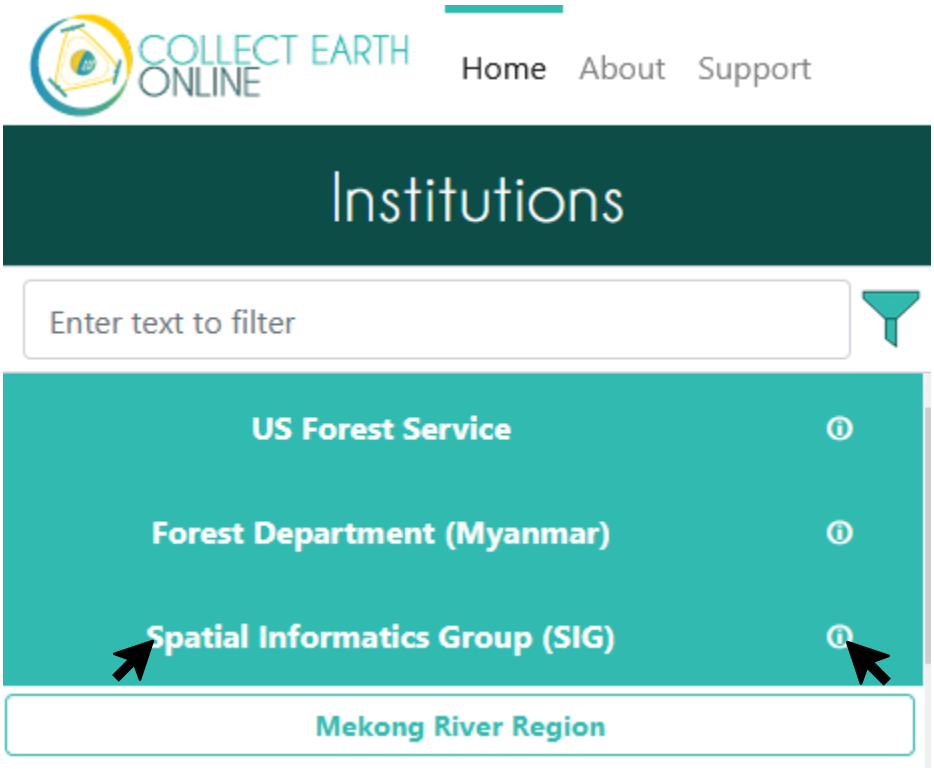
You will have to either log in or set up a GitHub account to log an issue. By logging in, the development team can contact you in case they need additional information to provide a solution to the issue or feature suggestion. Alternatively, if you are unable to register an account with GitHub, you can ask questions on the OpenForis forum, which is available here http://www.openforis.org/support.

1. When you have a GitHub account, navigate to the CEO GitHub Page.
2. To log an issue or request a new CEO feature, simply click on the green **[New issue]** button in the upper right-hand portion of the screen.
3. Type in a title that conveys the topic of the issue or request.
4. Then below, type in a detailed message summarizing the issue you have encountered or the additional functionality you would like to see in CEO.
5. Once you have finished providing the details, click the green **[Submit new issue]** button. This will log your request.

Part 2: Institution Set Up and Management

A. List of institutions & Institution pages

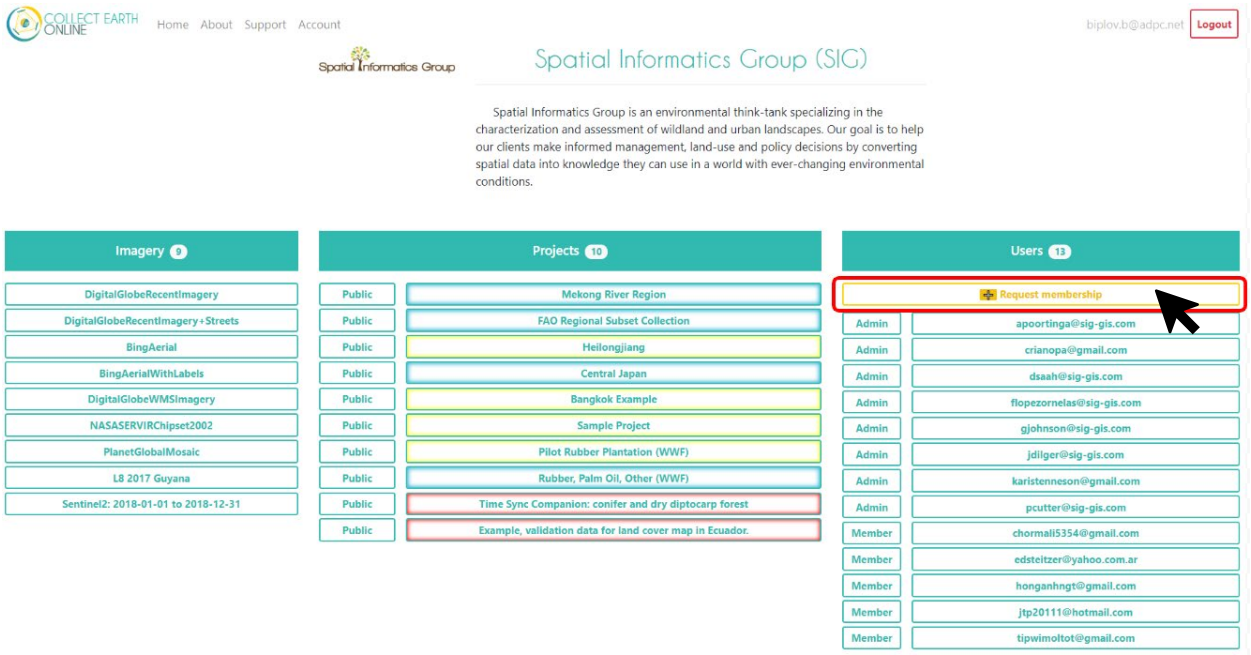
1. A list of all institutions is displayed on the **Home** page.
2. Clicking on an institutions' name expands a list of all active Projects for the Institution.



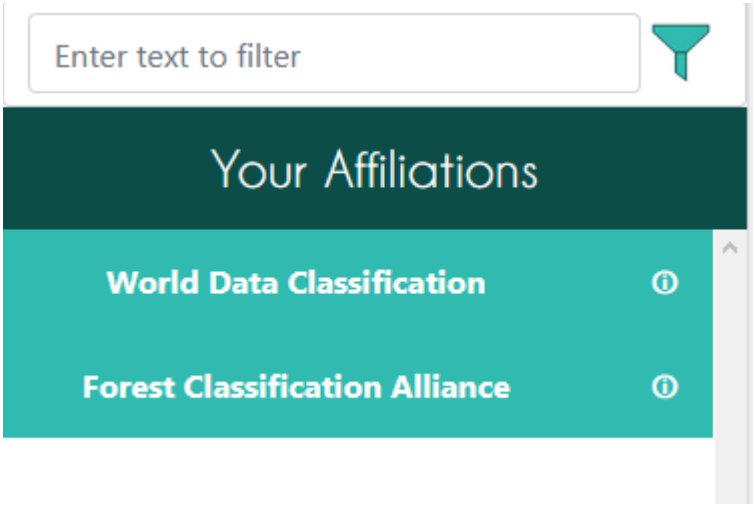
3. Next to each institution's name is an **info button**, symbolized by ⓘ.
4. By clicking on an institution's **[info button]**, the **Institution** page opens.
5. On any **Institution** info page, you can find:
   i. The institution's logo, which when clicked will take you to the institution's website link,
   ii. The institution's name
   iii. A short description of the institution,
   iv. Imagery feeds available to the institution
   v. The institution's projects, and
   vi. A list of all registered users associated with the institution.

B. Request to join an institution

1. If your institution already has an account on CEO and you would like to request to be a member, start by navigating to the **Institution** page via the **info button** described in A.2-4 above.
2. The list of Users is displayed on the right-hand panel.
3. Click on the **[Request Membership]** button to be considered to join the institution.

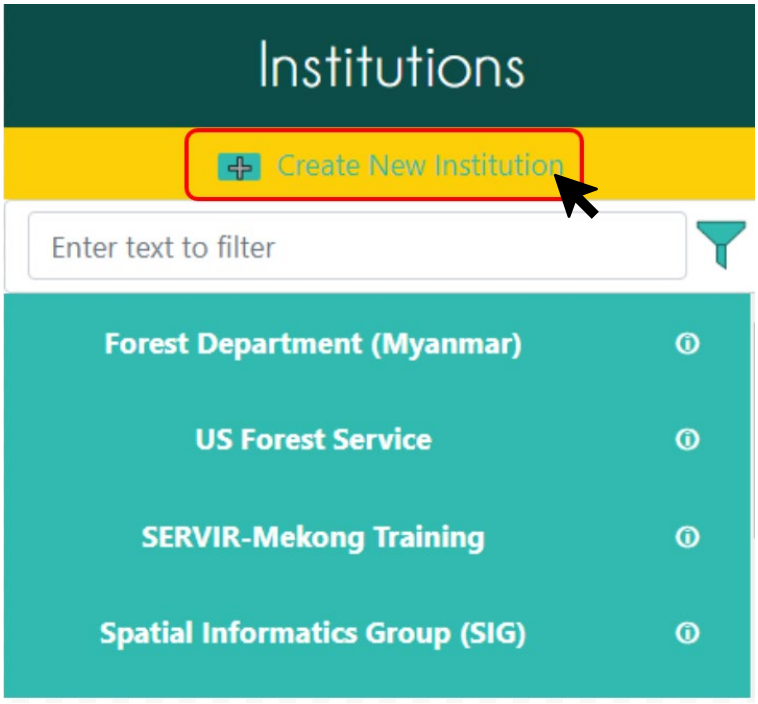


4. Once you are part of an institution, the **Institution** panel on the **Home** page will show your institutions. This makes it easier to access these institutions and their projects.

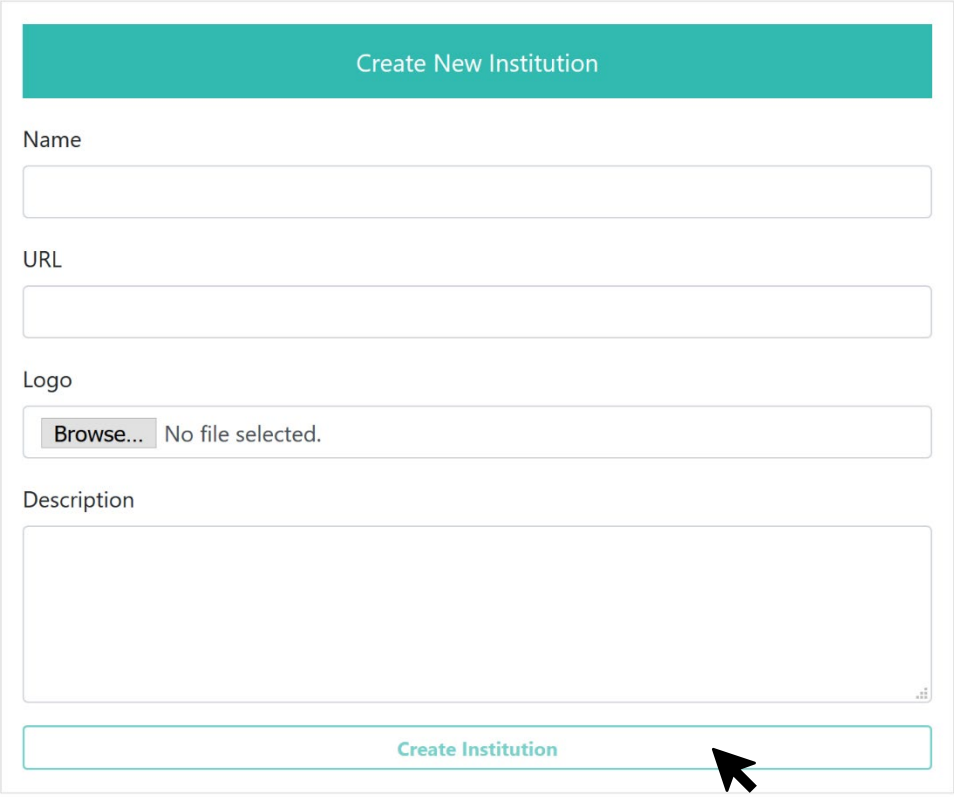


C. Creating a new institution

1. If your institution is not yet signed up for CEO, you can create a new institution. You will need to be signed into your account (click **[Login/Register]**).
2. When you are signed in, navigate to the **Home** page and click on **[Create New Institution]** on the left panel of the **Home** page.



3. Enter your institution's name, URL, and a brief description of your institution. These will display on your **Institution** page.
4. You can upload a logo from your computer by clicking on **[Browse…]** and navigating to the correct location on your computer.



5. Click **[Create Institution]** after you have finished entering your information.

D. Institution management

If you are an administrator for an institution, you can make changes to your institution's information after the institution is created.

1. If you want to make changes, start by navigating to your **Institution's** info page (see A.2-4 above).
2. You can make changes to the institution page by clicking **[Edit]** at the top of the page.
3. You can also delete your institution by clicking **[Delete]**. This action is PERMANENT, and your institution cannot be recovered afterwards. ALL PROJECTS ASSOCIATED WITH THE INSTITUTION WILL ALSO BE DELETED.

More importantly, you can also manage three aspects of your institution that allow data collection through CEO. These include the **Imagery** feeds, the data collection **Projects**, and the **Users** associated with the institution. This information is displayed in three panels on your **Institution's** info page.

4. The **Imagery** panel lists all available imagery and WMS Feeds. You can add new imagery feeds here as well. This is discussed in Part 3 of this manual.
5. The **Projects** panel lists your institution's projects, identifies projects as public or private, and allows you to create new projects. This is discussed in Part 4 of this manual.

6. The **Users** panel lists your institutions' members and allows them to be updated. This is discussed in E. below.

E. Institution user (member) management

1. As an administrator, you can add a Collect Earth Online member to the institution by typing the user's email address into the box and clicking the **[Add User]** button. Only email addresses that have already created a CEO account can be added to the institution. If you are successful, a pop-up window will notify you that the user has been given the role 'member,' and the user's email address will appear in the list of users.
2. You can approve pending affiliation requests as well.
3. To change the role of a user, use the menu to the right of their email address. Available user roles are Admin and Member. Users who are Admins can edit any of the institution's projects.
4. To restrict the admin capabilities of a user to only one of the institution's projects, you can create a new institution specifically for that project. Add the user as an Admin for the new institution.
5. You can also remove users from the institution using this drop-down menu and selecting the **Remove** option.

> A new user type, a "Project Reviewer" will be added in 2020. This change will come along with the planned Project Review functionality.

F. Other administrator privileges

1. As an administrator you can also review and modify the answers for all plots, including plots labeled by other CEO users. Members of an institution can only review and modify answers for plots they themselves have labeled.

Part 3: Built-in & Adding Base Imagery Sources

A. Built in base imagery sources

There are a handful of built in imagery options in CEO. Each of these options has different strengths and limitations.

Bing Aerial & Bing Aerial with Labels: The imagery provided by Bing Maps is composite satellite imagery. This means that each map tile is stitched together from imagery acquired on multiple dates. There is not a single date for an imagery tile. Some map tiles contain imagery collected over a multi-day window while other tiles contain imagery collected over a multi-year window. As there is not a single date for an imagery tile, CEO cannot provide the exact date of the imagery used. If you're interested in learning more, the Bing Maps API can be found here: https://docs.microsoft.com/en-us/bingmaps/rest-services/imagery/imagery-metadata.

> CEO is currently switching from the Aerial with Labels service, which is deprecated, to the Aerial with Labels on Demand service which replaces it.

The Bing tile system uses the Mercator projection and has 23 levels of zoom (though not all levels are available in all locations). Commonly, the resolution at max zoom is about 0.3 m per pixel. For more information, see https://docs.microsoft.com/en-us/bingmaps/articles/bing-maps-tile-system.

MapBox is an open source mapping platform for custom designed maps. The satellite imagery uses different sources based on the zoom level and geographic availability.

- **Zoom levels 0-8** use de-clouded data from NASA MODIS satellites.
- **Zoom levels 9-12** use NASA/USGS Landsat 5 & 7 imagery.
- **Zoom levels 13+** use a combination of open and proprietary sources, including Digital Globe for much of the world, USDA's NAIP 2011–2013 in the contiguous United States, and open aerial imagery from Denmark, Finland, and parts of Germany.

CEO has created a plain satellite layer as well as one with labels. More information is available here: https://www.mapbox.com/ and https://docs.mapbox.com/help/how-mapbox-works/satellite-imagery/.

> Maxar has deprecated multiple data products, including: DigitalGlobeRecentIMagery; DigitalGlobeRecentImagery+Streets; DigitalGlobeWMSImagery; and EarthWatch.

Landsat 5, 7, & 8, Sentinel 2, and data available through Google Earth Engine are available through the Geo-Dash interface (see Part 6 on how to add these options). However, these datasets will appear in a separate window and are not available as basemaps.

If you'd like, you can also pipe these through the GeoServer Imagery pathway described below.

B. Set up a Web Map Service (WMS) or Web Map Tile Service (WMTS) feed for basemap imagery

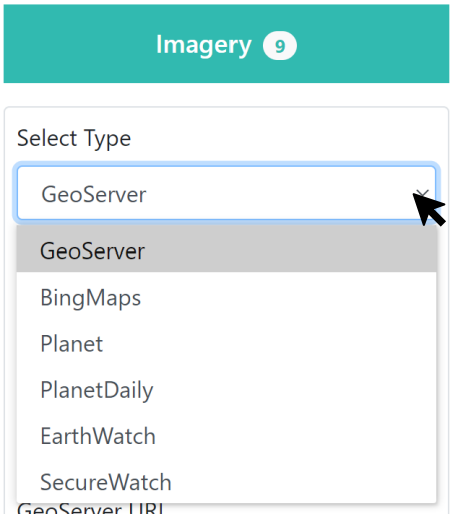
The built-in MapBox and Bing imagery are not sufficient as basemaps for some projects, including projects comparing land uses between two time periods, or projects that need imagery from specific dates or months. CEO allows institutions to add new imagery sources through their Institution page. Imagery must be accessible using WMS or WTMS. Images stored locally on a user's computer cannot be added to a CEO project, although they can be accessed if you upload them to a WMS or WTMS (as a basemap), or as a Google Earth Engine Asset (for Geo-Dash implementation).

The instructions below assume you are starting on your Institution page (see directions in Part 2: A.2-4) and are logged in as an Administrator for your Institution.

Under the **Imagery** panel on the **Institution** page, click **[Add New Imagery]**. You will see 6 options: GeoServer, BingMaps, Planet Monthly, Planet Daily, EarthWatch, and SecureWatch. GeoServer is a protocol that can be applied broadly, while the other options connect to specific data sources that you generally need to subscribe to.

For BingMaps, Planet Monthly, Planet Daily, EarthWatch and SecureWatch you will first need to locate your no-authentication API key. You will be asked to provide this key in the CEO imagery setup panel.

We will now talk about how to add imagery for each of these options, since the information needed to set up the WMS server will differ based on the specific imagery you are adding (note the options in the image may be different than what you see, as imagery availability has shifted significantly recently).



1. GeoServer
    i. GeoServer is an open-source Java-based software server that allows users to view and edit geospatial data. More information can be found here: http://geoserver.org/about/. This option can be used to access data your Institution hosts or data services that use this platform.
    ii. **Title:** This will be the displayed name of the imagery.
    iii. **Attribution**: This is the displayed attribution for your imagery. It will appear at the top of the map while users perform data collection tasks.
    iv. **GeoServer URL**: This should be the URL of your GeoServer. It should be http:// or https://
    v. **GeoServer Layer Name**: This is the layer name from your server that you want to display.
    vi. **GeoServer Params (as JSON object)**: If there are any parameters for the layer you want to specify, put them here as a JSON object. Documentation for the WMS getMapService is available online here: https://docs.geoserver.org/stable/en/user/services/wms/reference.html#getmap. These work only if they are parsed by the target GeoServer. Here is an example:

```
{"TILED": true}

{"VERSION": "1.1.1", "CONNECTID": "63f634af-fc31-4d81-9505-
b62b4701f8a9", "FEATUREPROFILE": "Accuracy_Profile",
"COVERAGE_CQL_FILTER": "(acquisition_date>'2012-01-
01')AND(acquisition_date<'2012-12-31')"}
```

> Quotes **MUST BE** "" ASCII/neutral/vertical quotation marks. Using “” smart/typographic/curved quotation marks will cause errors. Use only Unicode U+0022 and U+0027.

When all fields are filled out, click on **[Add New Imagery]**.

2. BingMaps
    i. This allows you to add Bing Maps with your own API key (instead of using CEO’s).
    ii. **Title:** This will be the displayed name of the imagery.
    iii. **Imagery Id:** Only Aerial and AerialWithLabels are currently implemented. Note that the AerialWithLables imagery uses the legacy static tile service, which is deprecated, and current data will not be refreshed. It therefore may have older imagery than the Bing Aerial dataset.
    iv. **Access Token:** Your BingMaps key. For more information or to obtain your own key, see https://docs.microsoft.com/en-us/bingmaps/getting-started/bing-maps-dev-center-help/getting-a-bing-maps-key.
    v. When all fields are filled out, click on **[Add New Imagery]**.

3. Planet Monthly
    i. Planet offers multiple data products (product specification here: https://support.planet.com/hc/en-us/articles/360022233473-Planet-Imagery-Product-Specifications) . This option pulls from the Planet Monthly mosaic product, which allows you to display imagery from a specific month.
    ii. **Title:** This will be the displayed name of the imagery.
    iii. **Default Year**: The default year that will be displayed when the map loads.
    iv. **Default Month**: The default month that will be displayed when the map loads. Use integer format 1-12.
    v. **Access Token**: Your Planet access token. For more information see https://developers.planet.com/docs/quickstart/getting-started/
    vi. When all fields are filled out, click on **[Add New Imagery]**.

> Default Year & Default Month will let you put in any integer, positive or negative. The up and down arrow keys start at 0. Please type year in the YYYY format and month as an integer between 1-12.

4. PlanetDaily
    i. PlanetDaily is another imagery product available from Planet. It allows users to detect land use and land cover change in near real time. This data source allows you to select a start and end date, with up to daily imagery resolution. Your study area might not have full coverage every day.
    ii. **Title:** This will be the displayed name of the imagery.
    iii. **Access Token:** Your PlanetDaily API key. For more information see https://developers.planet.com/docs/quickstart/getting-started/.
    iv. **Start Date:** Starting date for the imagery you are interested in; you can input the date using numeric keys or with the calendar widget on the right side.
    v. **End Date:** Ending date for the imagery you are interested in.
    vi. When all fields are filled out, click on **[Add New Imagery]**.

5. EarthWatch has been discontinued by Maxar.

6. SecureWatch Imagery
    i. SecureWatch is another service from Maxar focused on monitoring for new land use/land cover changes and comparing current land use/land cover with over 20 years of historic images. For more information see: https://www.digitalglobe.com/products/securewatch.
    ii. **Title:** This will be the displayed name of the imagery.
    iii. **Connect ID:** This is your API key. You need to use a no-auth key here, which should be a string of letters and numbers separated by dashes.
    iv. **Start Date:** Starting date for the imagery you are interested in; you can input the date using numeric keys or with the calendar widget on the right side.

v. **End Date:** Ending date for the imagery you are interested in.
vi. When all fields are filled out, click on **[Add New Imagery]**.

For PlanetDaily and SecureWatch imagery, the dates you input are the default dates that the imagery will be restricted to on the collection page. However, the user will be able to change these when exploring the map as there are start & end date widgets on the collection page sidebar (there are examples in the **Data Collection Manual**). For SecureWatch, the user will also be able to choose between FeatureProfiles. Without specifying a FeatureProfile, the most recent available imagery between the start and end dates displayed on the map.

EarthWatch and SecureWatch will not return imagery if the map is zoomed out too much. This results in a white map canvas being displayed at the project overview level usually. Simply click the "Go to first plot" button on the Collection page to zoom in to the plot level, and then the imagery should appear.

For SecureWatch, the date shown when data is collected will be added to the project .csv data available for download (See Part 7: E “Download your data”).

There is currently no way to ‘preview’ what imagery will be visible in your new layer in the ‘Add Imagery’ workflow. There are two ways to work around this.

First, if your data source has a data viewing portal, you can use this to explore the imagery and determine what is available for the time periods you are interested in.

Second, you can add the imagery layer, then open an existing project from your institution. The imagery will be available in the dropdown menu (if you are switching between a project window and an Institution window, you may need to refresh the project window to get the new layer to appear). You can then check if the imagery is displaying correctly and go back to the Institution page to try re-adding the imagery if your previous attempt failed.

C. Adding imagery from multiple time periods

Adding multiple imagery options with different default time periods can make data collection easier for projects that compare two or more time periods to detect land use and land cover change. WMS/WMTS that you can use to create basemaps from different time points include GeoServer, Planet Monthly, Planet Daily, and Secure Watch.

1. GeoServer:
    i. For GeoServer, how to add different years of imagery depends on your server.
        (a) If your different years are stored as different layers, alter the GeoServer Layer Name field when you add the second layer. Make sure your title/attribution/etc. fields are accurate for the new layer.
        (b) If your server uses filtering to display imagery from different years you will need to alter the GeoServerParams field (again, making sure the information in your other fields is correct).
    ii. Once you have decided the best approach for your server, repeat the steps in Part 3 B.1 above for each time period you would like to add.
    iii. **Note that some years may not contain any imagery**, due to the sparseness of the data within the database. If no imagery for the selected time range appears, you will need to change your GeoServerParams field, possibly to change the feature profile or date ranges.

2. Planet Monthly & Planet Daily
    i. For both Planet products, you simply need to change the time period fields to add layers with different default time periods. Users will be able to change the time period displayed during data collection; however, this is the default that will be shown first.
    ii. Be sure to change the Title field to reflect the correct default Year, Month, and Day for each new layer that you add.

3. EarthWatch & Secure Watch
    i. For both Maxar products, you simply need to change the time period fields to add layers with different default time periods. Users will be able to change the time period displayed during data collection; however, this is the default that will be shown first.
    ii. Be sure to change the Title field to reflect the correct default Year, Month, and Day for each new layer that you add.

> All imagery within your institution is available for all projects.
>
> If you want to limit which projects can use which imagery (e.g. to control costs), you will need to create a separate institution for that project and add your imagery source ONLY to that institution.

D. Estimating imagery costs

Before setting up a project in CEO, it is important to estimate how much imagery will be used for budgetary and resource allocation. Here is a quick guide to help.

1. When is imagery used? Imagery data is used whenever there is a map on the page. This means that on CEO, all these pages **can** use data:
    i. Home
    ii. Data Collection
    iii. Create Project
    iv. Review Project
    v. Project Dashboard
    vi. GeoDash (specific options or modules)
    vii. TimeSync (when implemented)

On these pages, when the map first loads, imagery data is used. Every time a user zooms or pans the map window, imagery is used. The largest amount of use will probably be with Data Collection.

2. Estimating imagery use for a project:
    i. Each organization that provides imagery sets their own rules for how many tiles you can download per year given the kind of account that you have with them. Therefore, it is important to estimate this before setting up a project.
    ii. Additionally, services may "count" imagery against your quota differently. For example, Planet uses a rule "*Every Pixel is Charged Once*" so you can download a pixel multiple times but it is only charged once (see here: https://support.planet.com/hc/en-us/articles/360021227554--When-is-a-Planet-product-charged-against-my-quota-). You will also need to know this.
    iii. To figure out how much imagery you are likely to use for a single project, count the number of plots. Then determine how many users will classify each plot (in CEO this is usually 1 user per plot currently). Next, try to factor in how often people will zoom or pan their maps for context when answering the survey questions for a plot. Multiply these numbers together.
    iv. Next, decide if you are using the Geo-Dash and ask yourself how many map widgets you will display on your Geo-Dash page. Multiply that number by the number of plots to get the amount of Geo-Dash imagery you will need. Keep in mind that Geo-Dash imagery counts against our annual limit for user memory/processing in GEE, whereas the usual global layers on CEO (EarthWatch, Bing Maps, SecureWatch, Planet) have separate annual tile-based limits. For GEE, we recommend clipping and pre-processing the imagery to image assets or imageCollection assets for the collection area. This eliminates processing on-the-fly for each user that is collecting, as Geo-Dash can just grab the pre-processed image asset.
    v. Finally, add a few extra tile downloads for loading the maps while creating and reviewing the project.

vi. Once you have a sense of how many map images you will need for your project, you will then need to look up the tile counting policy for the imagery service that you are using. For example, some of them count 15 tiles as 1 unit of usage. Others use different counting rules.

3. Tips to reduce imagery use: Consider setting your default background imagery to a cheaper source—like Bing—and only switching to more expensive paid imagery when you are at the correct zoom level.

Part 4: Project Creation

With your imagery loaded, you are ready to create a new project.

Before starting a data collection effort in CEO, make sure you have concrete goals, indicators, and metrics for your initiative. While this manual focuses on technical issues, other resources are available to help you create these goals, indicators, and metrics. Some helpful resources include:

- The *Road to Restoration: A Guide to Identifying Priorities and Indicators for Monitoring Forest and Landscape Restoration*, found at: http://www.fao.org/in-action/forest-landscape-restoration-mechanism/resources/detail/en/c/1253837/. This guide from FAO & WRI outlines steps toward setting goals, choosing indicators, and defining metrics.

If your project is focused on land use or land cover classification, you also need to have a classification scheme and an interpretation key. A classification scheme should be exhaustive, exclusive, consistent with the purpose of the work, and sufficiently descriptive. The scheme can be either single level or hierarchical and it does not need to be of uniform detail. Data collectors refer to an interpretation key—a collection of rules, imagery, and guidance—to classify land cover elements. A comprehensive interpretation key allows data collectors to produce more consistent and reliable results.

- See Chapters 2 & 3 of the Theoretical Manual from Collect Earth Online, found at https://collect.earth/downloads/CEO_Theoretical_Manual.pdf.

Designing a project is an iterative process, and you will probably need to make multiple projects in CEO as you refine your goals, land use or land cover classification, imagery sources, etc. Some things can be changed after a project is published, while others currently cannot. For example, if you realize after you have created a project that you have forgotten to add an imagery layer, you can add it after project setup with the steps from Part 3.

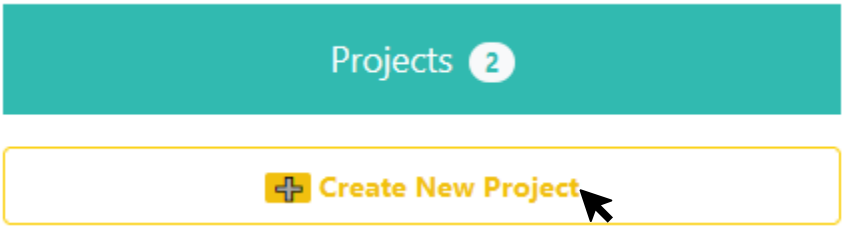
> Project review functionality will be coming soon to CEO. This functionality will allow users to create a project, perform test data collection, then refine the project and project questions before publication. Each time the questions are changed, all test data will be erased. This is done to make sure that project answers are consistent.

The survey questions that are asked about each survey point cannot currently be changed after the project is published. Do not start data collection until you are sure that your survey questions are correct. The CEO team is currently working on a 'Preview' mode that allows administrators to iterate through survey question creation and data collection before publishing the project.

The instructions below assume you are starting on your **Institution** page (see directions in Part 2: A.2-4) and are logged in as an Administrator for your Institution.

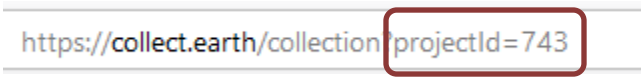
A. Create a new project

1. In the **Projects** column on your **Institution** page, click [Create New Project]. This will bring you to the Create Project form.

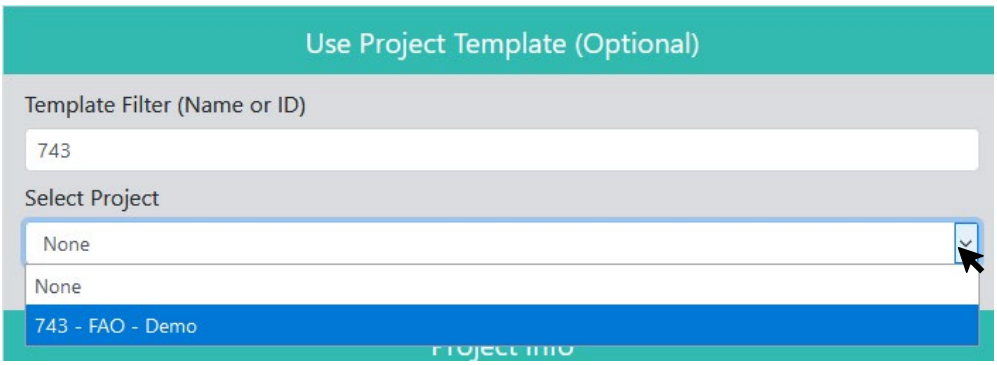


B. Use project template (optional)

1. This section is used to copy all the information—including project info, area, and sampling design—from an existing published project to a new project. This is useful if you have an existing project you want to duplicate for another year or location, or if you are iterating through project design. For a template, you may use any available published or closed project from your institution. You cannot use deleted projects. You cannot use another institution's private project (only members or admins can access), but you can use another institution's public project (any website visitor or CEO user). For more on project privacy settings, see below.
2. If you do not want to copy another project, simply skip this section, and leave the **Select Project** field set to **None**.
3. **Template Filter (Name or ID):** Type in a keyword in an existing project's name or the Project ID Number. You can find the Project ID number by navigating to the project you want to copy and looking at the URL.



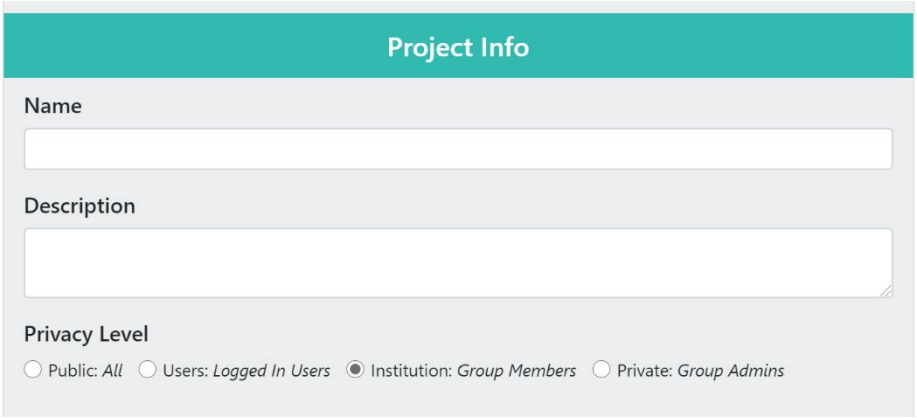
4. Then click on the dropdown menu under **Select Project** and click on the project's name



If you select a project under **Select Project**, and then if you change **Select Project** back to **None** the Project AOI will remain on the map, and the basemap may display only gray/white. Change the **Basemap Source** to any other source and then back to the source you are interested in and the basemap will reappear.

5. This will add two checkboxes, **Copy Template Plots and Samples** and **Copy Template Widgets**. They are checked by default. If **Copy Template Plots and Samples** is checked, the Plot Review and Sample Design sections will only display an overview of the number of plots, etc. Uncheck this box to change those parameters. **Copy Template Widgets** refers to Geo-Dash options covered in Part 6: Geo-Dash Implementation.

C. Project info



1. Under **Project Info**, enter the project's **Name** and **Description**. The **Name** should be short and will be displayed on the **Home** page as well as the project's **Data Collection** page. You should keep the **Description** short but informative. Users will see these if they click on the project's pin on the map on the home page. You will also see this when you are administering your project.
2. If you are using a template, the **Name** and **Description** will automatically be populated. Be sure to change this to reflect your new project.
3. The **Privacy Level** radio button changes who can view your project, contribute to data collection, and whether admins from your institution or others creating new projects can use your project as a template.
   i. **Public: *All*:** All users can see and contribute data to your project. Admins can use your project as a template.
   ii. **Users: *Logged in Users*:** Any user logged into CEO can see and contribute to your project. Admins can use your project as a template.
   iii. **Institution: *Group Members*:** Members of your institution can see and contribute to your project. Admins from other institutions cannot use your project as a template.
   iv. **Private: *Group Admins*:** Only your Institution's Admins can see and contribute to your project. Admins from other institutions cannot use your project as a template.

D. Project AOI: Setting the project area of interest

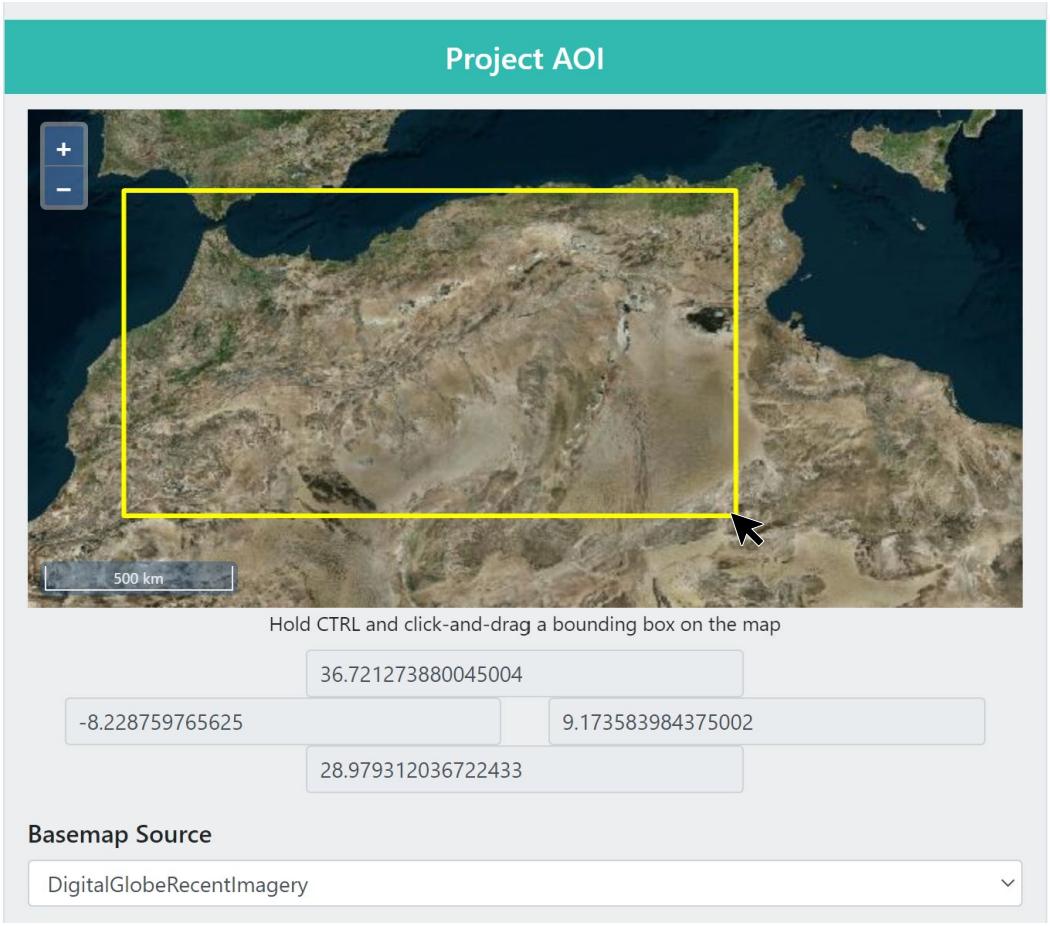
The project area of interest (AOI) determines where sample plots will be drawn from for your project. This is the first step in specifying a sampling design for your project.

There are two main approaches for specifying an AOI and sampling design: 1. using CEO's built in system and 2. Creating a sample in another program (QGIS, ArcGIS, etc.) and importing it into CEO. If you are interested in uploading your own sample as a .csv or .shp file, please see Part 4: H Uploading CSV & SHP files.

1. The easiest way to select your project AOI is by drawing a box in the map window.

i. Locate your area of interest by zooming in/out using the scroll wheel of your mouse, or the + and – boxes in the map window. You can pan the map by clicking on it and dragging the map window.
ii. Hold the CRTL-key (command key on a Mac) down and draw a box while keeping the left mouse key pressed down.
iii. The coordinate boxes will populate once the box is drawn and you let your mouse key go. Coordinates are displayed in lat/long using **WGS84 EPSG:4326**.

> You cannot upload a shapefile (.shp) to delineate your AOI only.
>
> Instead, use ArcGIS, QGIS, or a similar program to create sample points within your desired AOI.



2. You can change the **Basemap Source**, which specifies the default imagery that users will see when they begin data collection on your project. The map view will display the current selection. Your users can switch between all the available imagery layers during analysis.

3. The default options are MapBox, Bing Aerial, and Bing Aerial with Labels on Demand.
4. You may have more options based on what imagery your institution has added. You can choose all public and private WMS feeds from your (& the project's) institution.

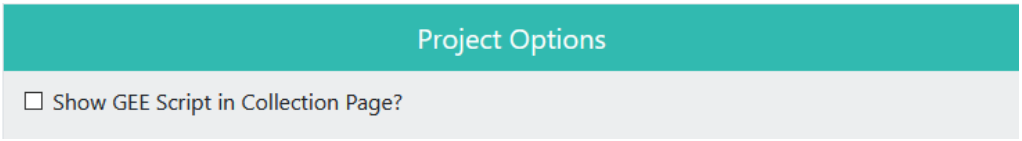
> Maxar has deprecated multiple data products, including: DigitalGlobeRecentIMagery; DigitalGlobeRecentImagery+Streets; DigitalGlobeWMSImagery; and EarthWatch.

PlanetMonthly, PlanetDaily, EarthWatch, and SecureWatch do not allow for large area data pulls, so it should not be your default basemap (users will just see a white box). You will need to set a different default basemap and have your data collectors switch to PlanetDaily once they have zoomed in on a plot to interpret.

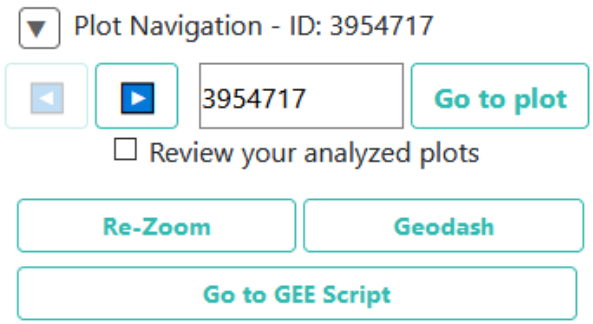
5. If your project is comparing land use and cover changes between two years, select one of your focal years' WMS imagery as the default imagery here. Your users can then easily switch between this year's imagery and the other year's imagery in data collection.
6. You will need to first set up the imagery feed for one date period under the institution imagery management panel. Refer to instructions in Part 4, Section B.

E. Project options

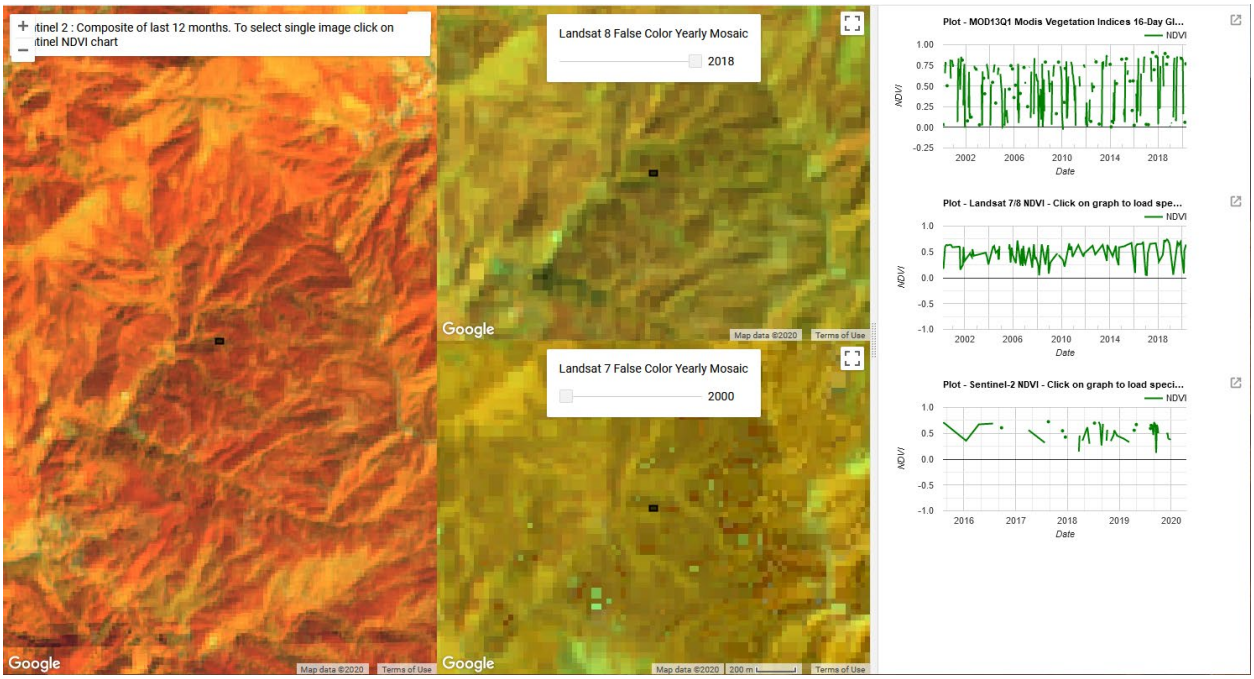
1. There is one option under Project Options, which is to show a GEE Script on the Collection page.



2. This allows users in Data Collection to click on a button labeled **Go to GEE Script**.



This button will take them to a website with additional information about the plot. An example is shown here:



F. Sample plot design

A well-prepared sample can provide a robust estimate of the parameters of interest for the population (percent forest cover, for example). This is often the most important factor in producing a reliable inventory or accuracy assessment. The goal of a sample is to provide an unbiased estimate of some population measure (e.g. proportion of area), with the smallest variance possible, given constraints including resource availability.

The first step of the built in CEO sample design is the plot design. The second step, selecting samples within a plot, is covered in the next section (Part 4: G Sample (point) design). Using these two steps, many sampling designs are possible with CEO. For information on uploading your own sample as a .csv or .shp file, please see Part 4: Uploading CSV & SHP.

For example, suppose you are quantifying forest cover on a landscape. For this approach, sample points are used to classify land cover and are then summarized at the plot level to create an estimate of the plot's forest cover percent. Information about plots can then be used to estimate the forest cover at the landscape level and detect patterns or trends. The accuracy of your landscape level estimates will depend on how many plots you classify & how variable the landscape is, among other factors. More detailed theoretical information is available in CEO's Project Development Manual (found at https://collect.earth/downloads/CEO_Theoretical_Manual.pdf).

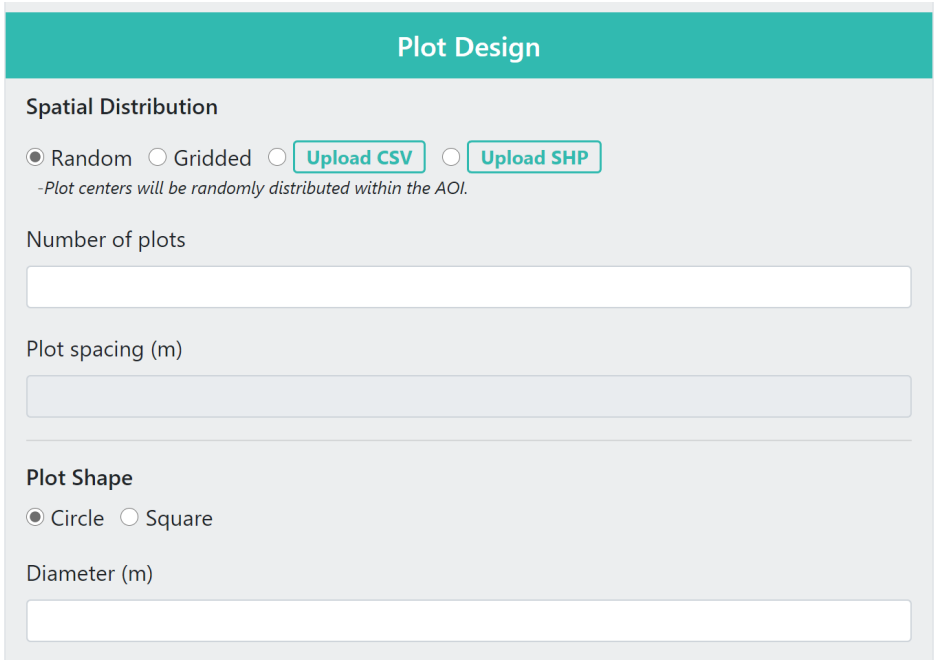
Also, some terminology might help for the two types of sampling available in CEO. **Simple random sampling** means that all points have an equal probability of being selected. It produces unbiased

parameters. However geographic balance with small sample sizes can be difficult. Also, rare classes may not receive sufficient coverage. If you have rare classes you want to detect, we highly recommend the stratified sampling approach from SEPAL discussed yesterday.

The second type is **systematic gridded sampling**. This is a grid of points placed over the landscape at regular intervals. This provides excellent geographic balance, but it is not possible to calculate unbiased estimates of population metric variance.

In the **Plot Design** section, you can specify the type and number of sample plots.

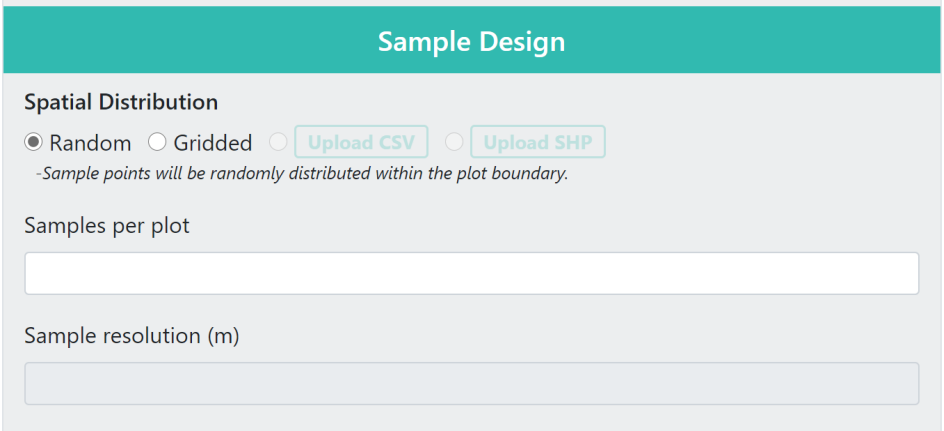
1. **Spatial Distribution** defines the distribution of the sample points. In CEO, you can specify either a random or a gridded (spatial systematic) sampling approach.
    i. Random sampling has the advantage of being extremely simple and producing unbiased parameters that are straightforward to calculate. However, geographic balance is not certain with smaller sample sizes, and rare classes may not be adequately sampled unless the sample size is large.
    ii. Systematic sampling has the advantage of providing excellent geographic balance. However, it is not possible to calculate a truly unbiased estimate of the variance of population metrics when using systematic sampling. Additionally, if patterns in your landscape match up with the spacing of your systematic gridded points, you will produce a very biased estimate.
    iii. If you select **Random**, you will need to provide the number of plots for the whole project.
    iv. If you select **Gridded**, you will need to provide the spacing between the centers of the plots (in meters).
    v. The maximum number of plots for a project is 5,000. For gridded sampling, you may need to increase the space between plots to avoid exceeding 5,000 plots.
2. **Plot shape** can be either a Circle or a Square. You will need to specify the plot diameter (for a circle) or the plot width (for a square). These sizes should be driven by the needs of your project. If they are small, your users will need to zoom out significantly to see the relevant background imagery because CEO automatically centers and zooms in to the plot's boundaries.


Plot Design
Spatial Distribution
Random
Gridded
Upload CSV
Upload SHP
-Plot centers will be randomly distributed within the AOI.
Number of plots
Plot spacing (m)
Plot Shape
Circle
Square
Diameter (m)

G. Sample (point) design

The second step is to determine how many sample points are within each plot, and whether they are sampled using random sampling or gridded sampling.

1. Under **Spatial Distribution**, with **Random** sampling sample points will be randomly distributed within the plot boundary. You will also need to specify the number of **Samples per plot**.
2. With **Gridded** sampling, sample points will be arranged on a grid within the plot boundary. You will need to specify the distance between points within the plot under **Sample resolution (m)**.
3. With **Center** a sample point will be placed in the center of the plot; you do not need to specify anything else.
4. The maximum number of sample points per plot is 200.
5. The maximum total number of sample points for the project (number of plots times the number of points/plot) is 50,000.



H. Uploading CSV & SHP files

While the default sampling design will work for many users, you may want to create your own sampling design and upload it to CEO using the .csv or .shp file capability. For example, this functionality is useful when: you want to draw your sample plots from within a shape other than a rectangle (e.g. the outline of a region or country) or if you want to do any stratification in your sampling design. You can create a .csv or .shp with your desired sampling plots/points through services and applications including Sepal, ArcGIS (e.g. https://pro.arcgis.com/en/pro-app/tool-reference/data-management/create-random-points.htm), and QGIS (e.g. https://freegistutorial.com/how-to-create-random-points-inside-polygon-on-qgis/).

You must use WGS84 EPSG:4326 format for coordinates in both .csv and .shp files.

For .csv files, specify plot centers by uploading a .csv with these columns: LON, LAT, PLOTID.

You can also upload a second file where you specify your own sample point centers by uploading a .csv with these columns: LON, LAT, PLOTID, SAMPLEID. LON and LAT can also be LONGITUDE and LATITUDE. You can have additional columns with data about your sample plots and points if and only if they come after these key fields.

If you do not specify the column names correctly (spelling or order), you will get the following error:

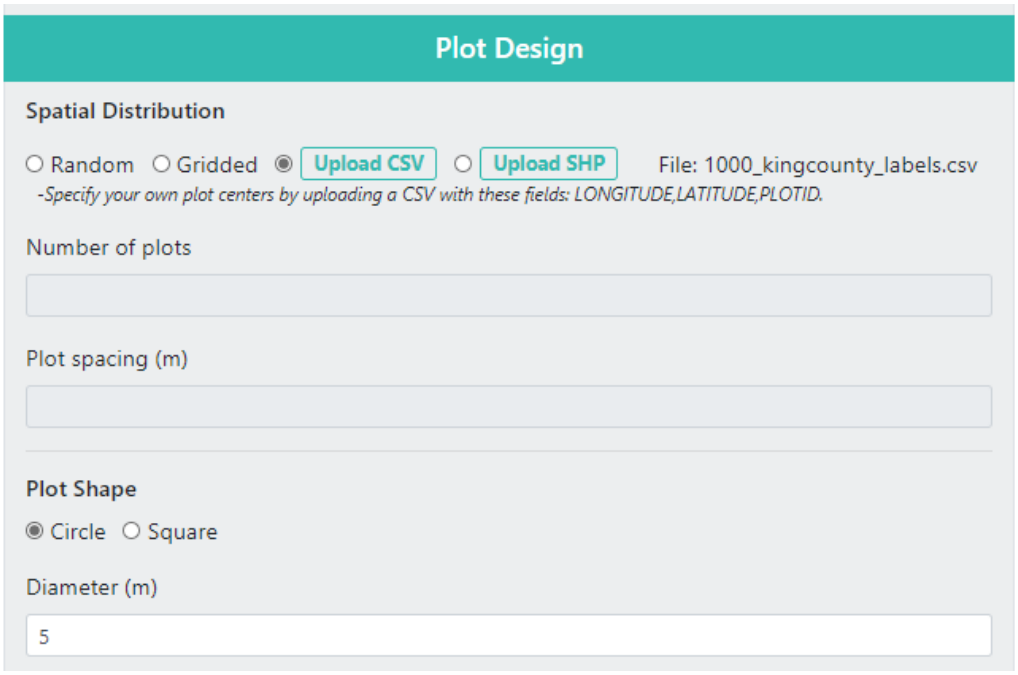
Longitude should be between -180 and 180, while latitude should be between -90 and 90. If you mix them up, you may get an error if your longitude is greater than 90 or less than -90 (when this is mixed up with latitude, it is 'above' the pole). Double check these values.

You can upload just one file for the plot centers OR two files, one for the plot centers and one for the point centers. When your .csv files are fit the above specifications, follow the directions below.

For one file with just the plot centers:

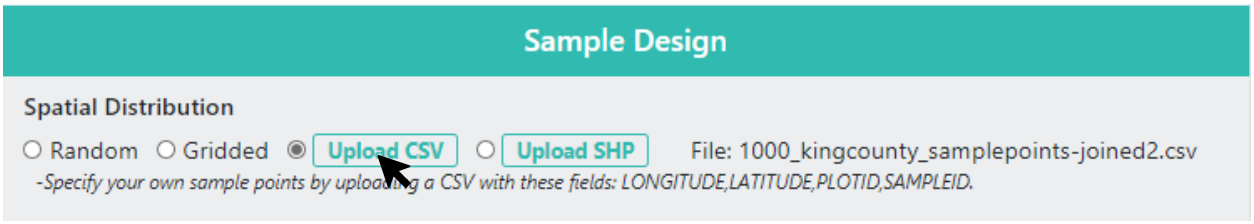
1. Under **Plot Design**, click on the radio button next to **Upload CSV**. You must have the radio point selected before the button to upload becomes available.
2. Then, click on **[Upload CSV]** and navigate to the .csv with your sample plot centers.

3. After you upload the file, the file name will appear next to "File:".
4. You will need to specify the **Plot Shape** and corresponding **Diameter** or **Width** (See Part 4: E Project Options above).
5. You will also need to specify your **Sample (Point) Design**, as in Part 4: F Sample Plot Design above.



For two files, one with plot centers and one with sample point centers,

6. Follow steps 1-4 above.
7. In step 4, it is important that you specify a plot size that is large enough to contain your points if they are also uploaded through a .csv or .shp. This will not be a problem if you use the built-in sample point design function.
8. Then, under **Sample (Point) Design**, click on the radio button next to **Upload CSV.**
9. Click on **[Upload CSV]** and navigate to your .csv
10. This will also work with **Upload SHP**.

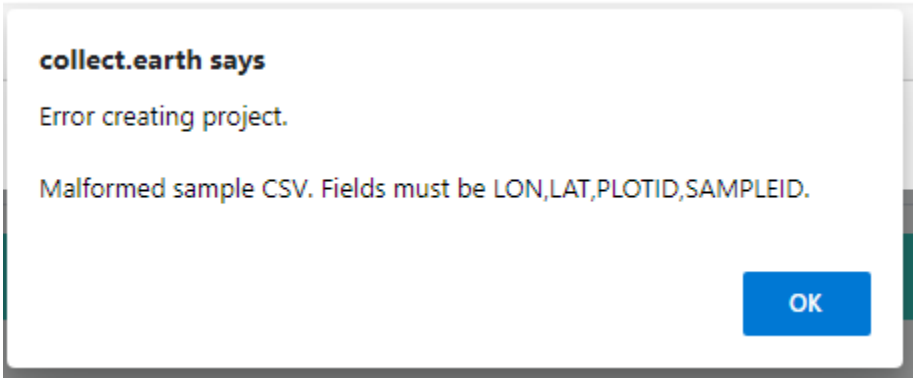


For .shp files, you can specify your own plot boundaries by uploading a zipped Shapefile (containing SHP, SHX, DBF, and PRJ files) of polygon features. Each feature must have a unique PLOTID field. LON and LAT are not required for polygons.

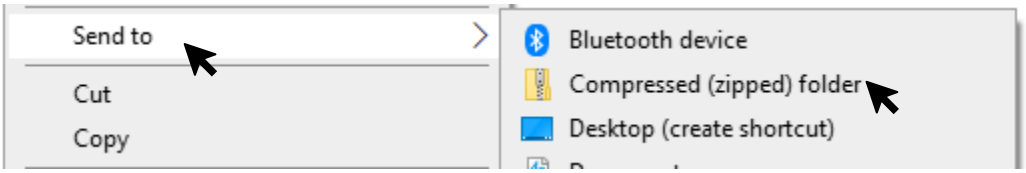
You will also need to upload a second file where you specify your own sample points. This can be a .csv with these columns: LON, LAT, PLOTID, SAMPLEID. LON and LAT can also be LONGITUDE and LATITUDE.

It can also be a zipped .shp file (containing SHP, SHX, DBF, and PRJ files). Each feature must have PLOTID and SAMPLEID fields. Either points or polygons will work for the sample point file, though point files must include LON and LAT. As with .csv files, you can have additional fields with information about your plots and points if and only if they come after these key fields.

If you do not specify your PLOTID in the sample point .csv or .shp zip file, you will get the following errors:



You can zip your files easily in Windows by selecting the relevant files, right clicking on one, and the clicking **[Send to] -> [Compressed (zipped) folder]**.



For a Mac, select the relevant files, right click on one of the files, and select **[Compress Items]** from the pop-up menu.

When your .shp files fit the above specifications, follow the directions below.

11. Under **Plot Design**, click on the radio button next to **Upload SHP**. You must have the radio point selected before the button to upload becomes available.
12. Then, click on **[Upload SHP]** and navigate to your zipped .shp file.

13. Then, under **Sample (Point) Design**, click on the radio button next to **Upload SHP**.
14. Click on **[Upload SHP]** and navigate to the zipped .shp file with your sample point points or polygons and select it.
15. This will also work with **Upload CSV**.

When you download your collected data, any column with extra information that were present in the uploaded .csv and .shp files will be preserved in the downloaded data (See Part 6: B).

CEO is currently developing a feature where any extra columns with additional data about your sample plots and points will be displayed in a small table during Data Collection.

I. Survey design—simple interface

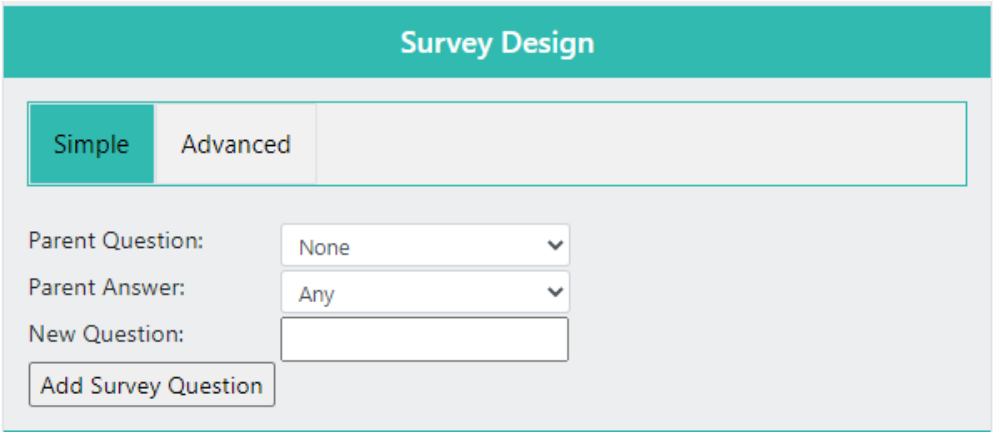
This is where you design the questions that your data collectors/photo interpreters will answer for each of your survey plots. Each question creates a column of data. This raw data facilitates calculating key metrics and indicators and contributes to fulfilling your project goals.

Survey Cards are the basic unit of organization. Each survey card creates a page of questions on the Data Collection interface.

There are two interfaces for **Survey Design.** For both interfaces, the basic workflow is: Create new top-level question (new survey card) → populate answers → create any child questions & answers → move to next top-level question (new survey card) & repeat until all questions have been asked.

We will now go into more detail about the two interfaces, the types of questions that can be asked, and when these questions might be useful.

The **Simple** interface provides a straightforward way to ask **multiple-choice** questions. In CEO, these questions are called "button—text" questions as in data collection they display as a button with text.



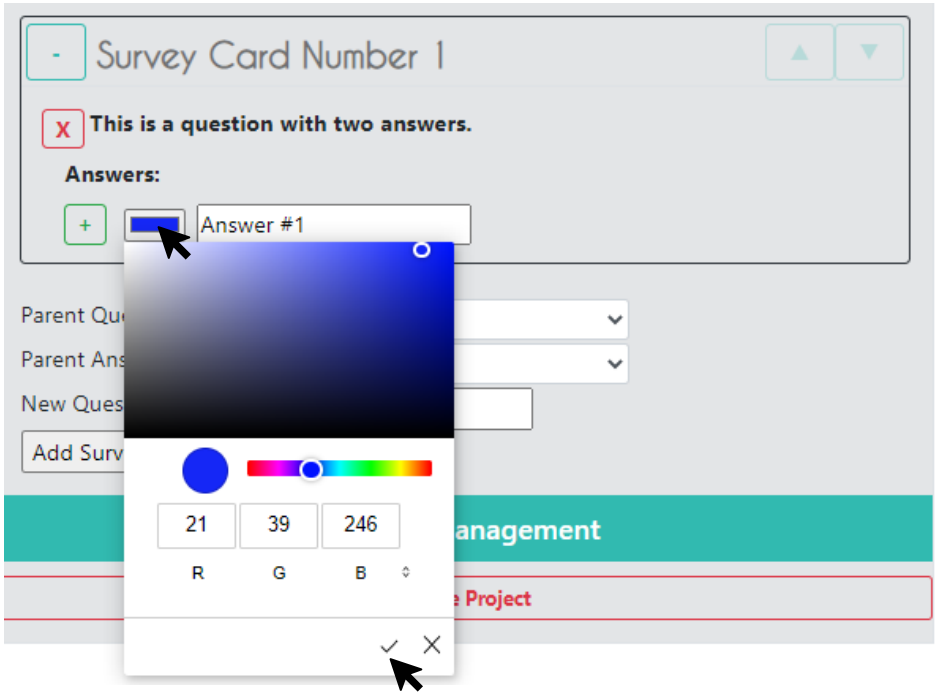
Questions of this type are useful for land use and land cover surveys, or anywhere where you want the user to choose between a limited set of mutually exclusive options.

You can also ask follow-up questions based on a user's response to further refine information about the plot. For example, if a user categorizes a plot as forest, you can follow up by asking them if it is deciduous or coniferous forest.

1. To start, type your first question into the **New Question** box. Since it is your first question, you cannot assign a **Parent Question** or **Parent Answer** for the question.
2. Try to keep the question text below 45 characters so the whole question will be displayed during data collection.
3. Click **[Add Survey Question]** to create your first Survey Card.
4. You can now add **Answers** to your question. **Answers** have two parts, a color and a text box.

i. Click on the **[blue rectangle]** to bring up the **Color Selector.** You can move the color selector dot or type in RGB values (0-255). Click the gray **[✔]** when you are done. The color you choose will be associated with the answer. When a data collector selects that answer, any sample points assigned that answer would also be assigned that color to display on the map.

ii. You can type your answer into the **Text Box**. Try to type answers with around 15 characters or less so that the full name is displayed during data collection.



5. Click the green **[+]** symbol to add the answer.
6. Continue adding answers until all the answers to your survey question have been added.
7. Now that you have a top-level, or parent, question with answers, you can add child questions and child questions that appear only when specific answers are chosen (have parent answers).

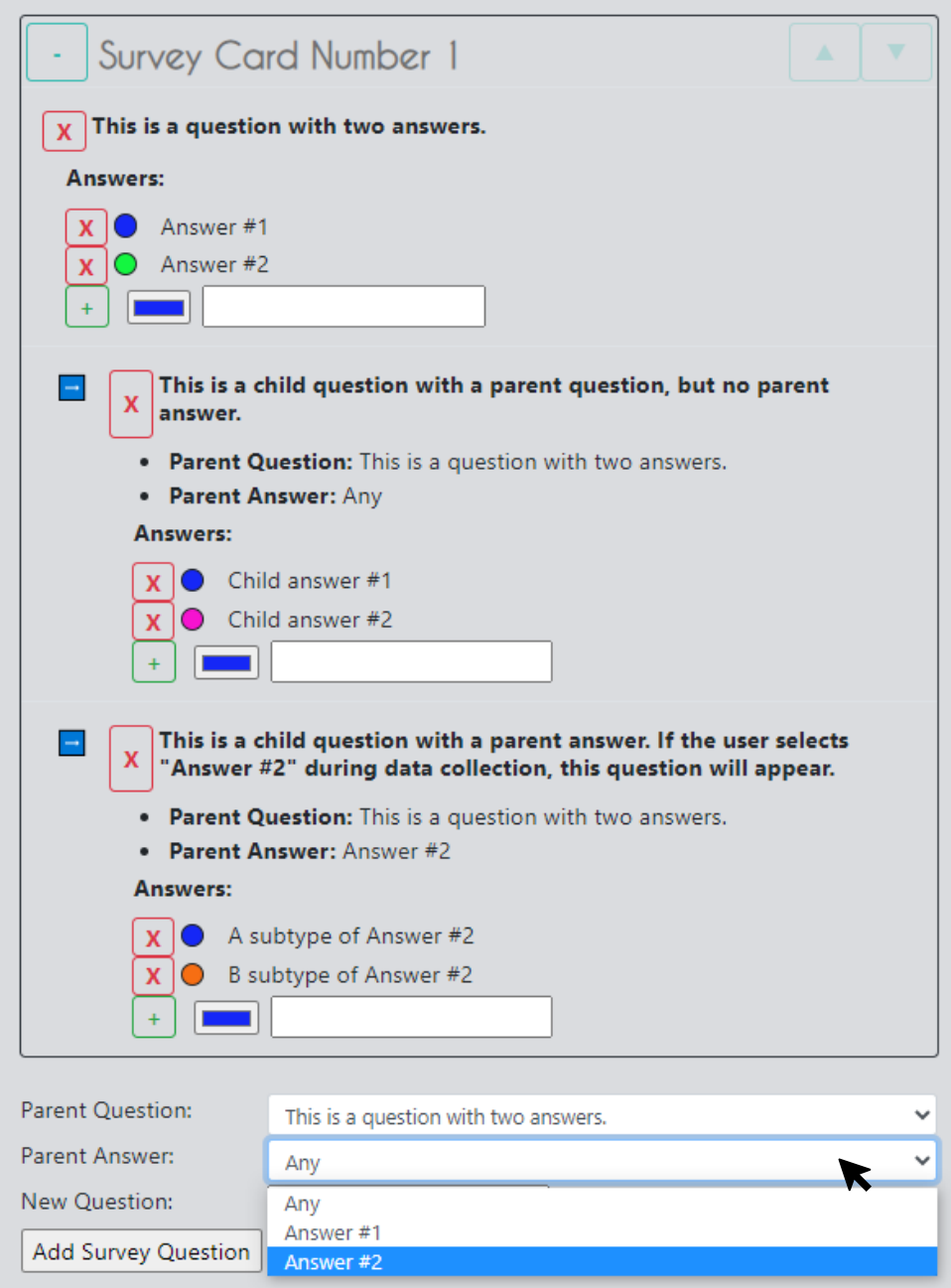
Parent & child questions, particularly child questions that have parent answers, are useful when you have broad categories and then want to refine the answer within that category.

8. To create a child question, next to **Parent Question** select the question you want.
9. You can then assign a **Parent Answer** through the dropdown menu. When this answer is chosen, the child question will appear.
10. If you do not want to assign a **Parent Answer**, then set the **Parent Answer** field to **Any.**
11. Once you have finished adding child questions with their answers, you can create a second survey card by setting the **Parent Question** field to "**None**" and the **Parent Answer** field to **Any**.
12. You can delete a question by clicking the red **[X].**

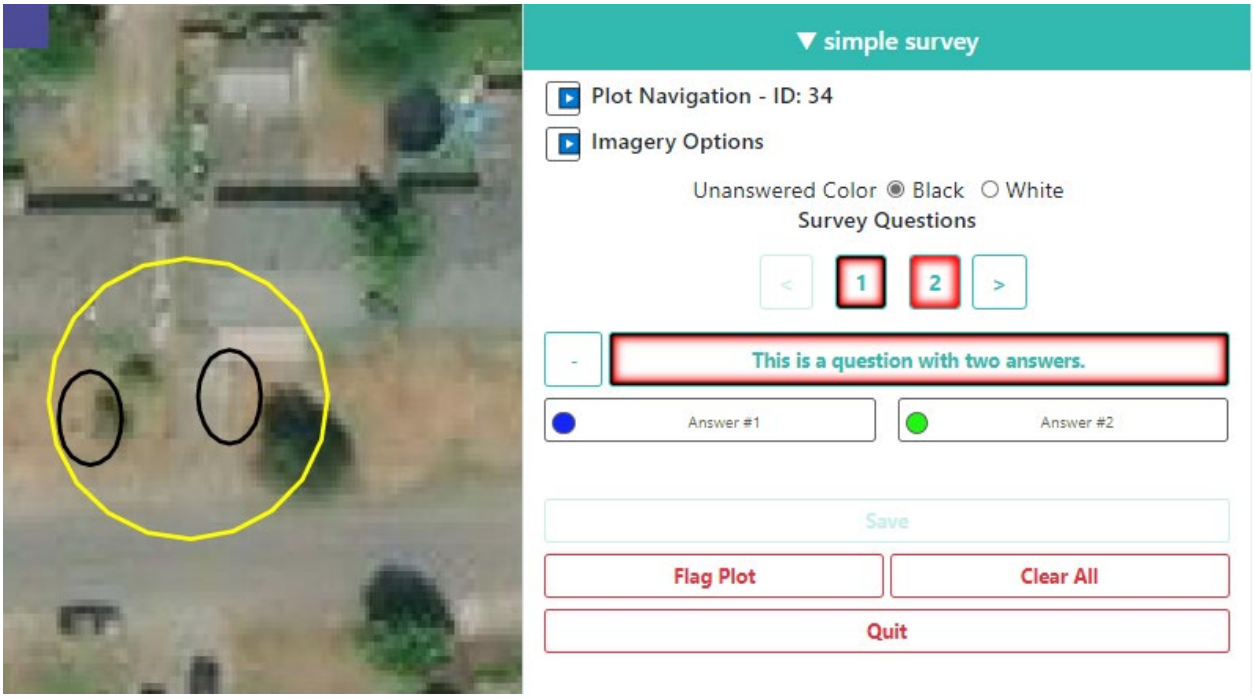
13. Deleting a parent question with children will delete the children questions as well.
14. You can collapse a survey card by clicking the **[-]** symbol in the upper left.

> You can change the order of the Survey Cards by clicking the blue up & down arrows in the top right.
>
> However, **DO NOT DO THIS**. Pressing these buttons will crash the project and you will lose all your progress.

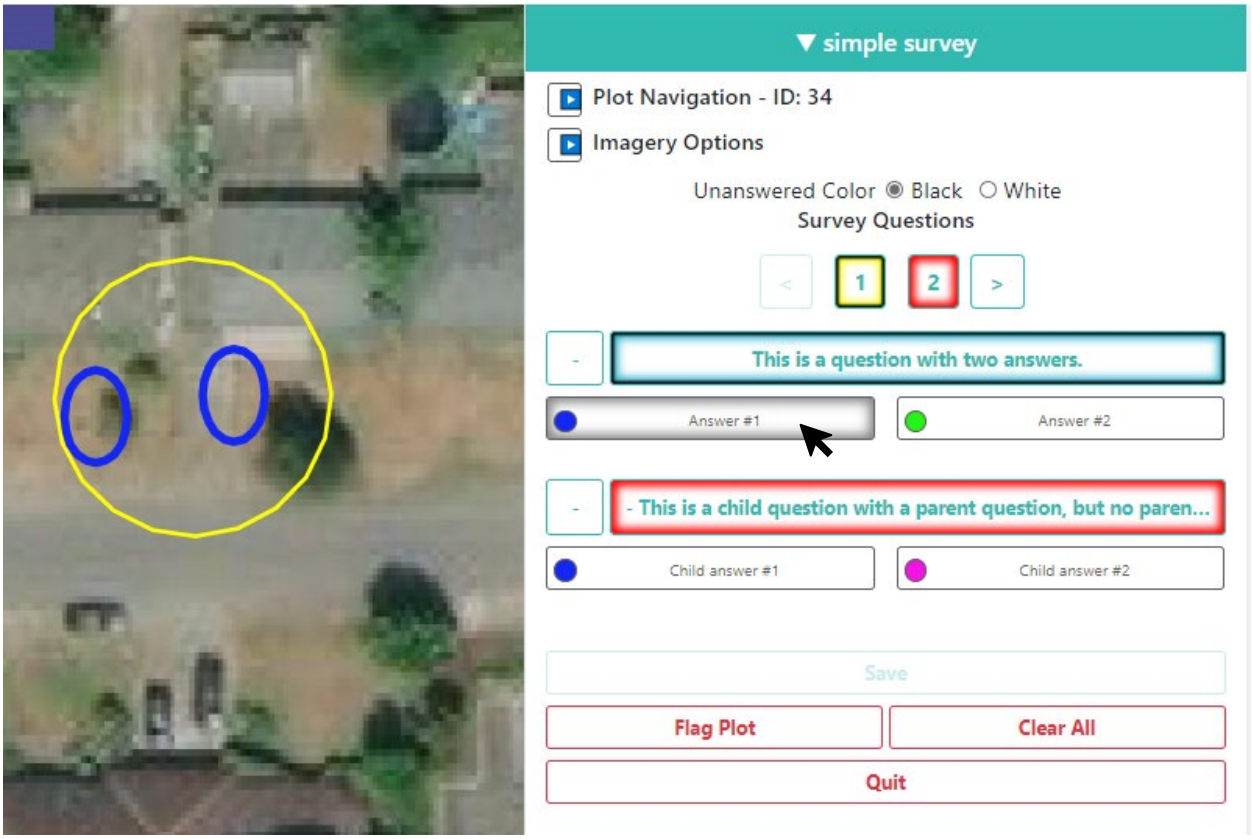
Here is an example Survey Card:



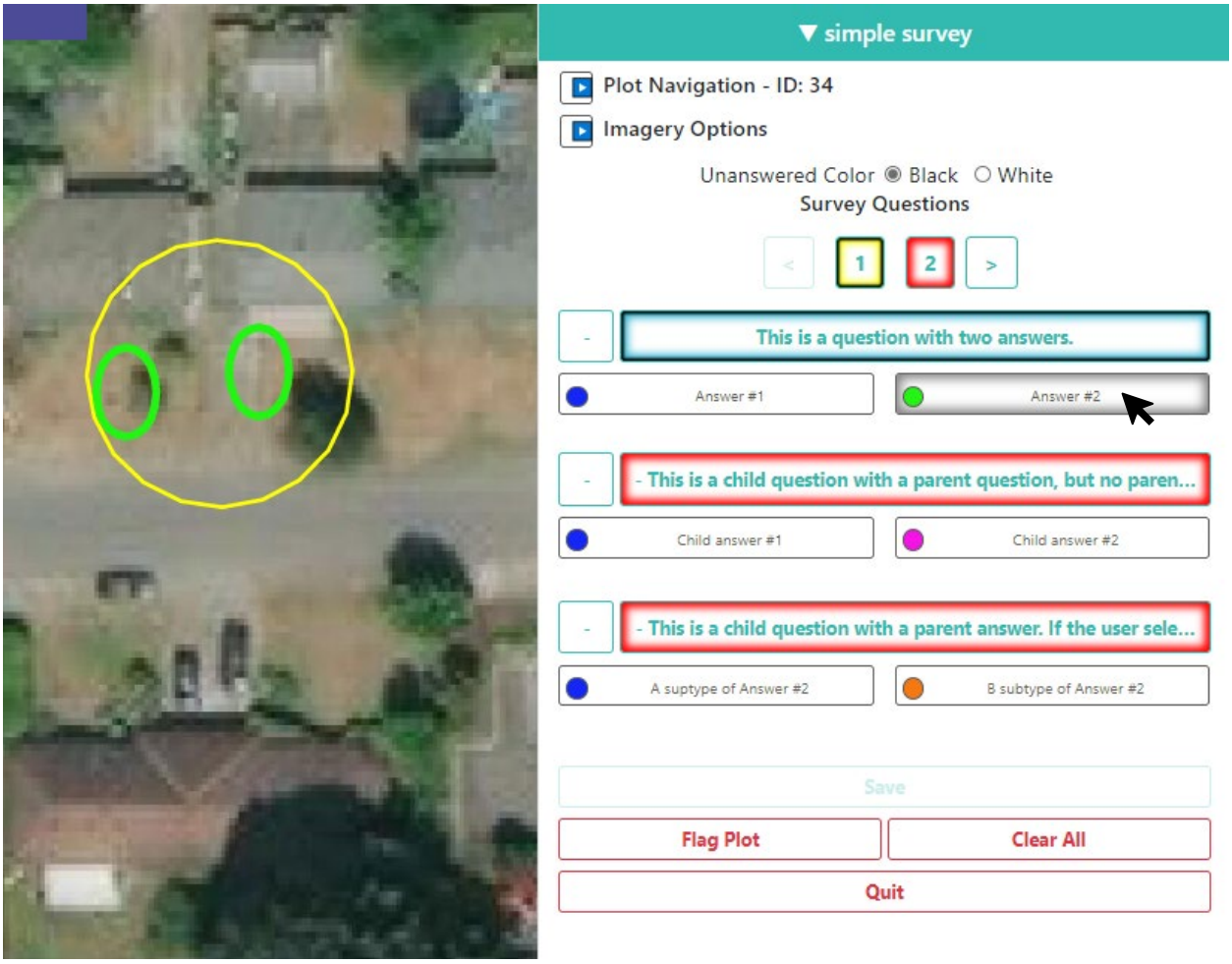
And here is how that survey card appears on the data collection page initially:



And after selecting “Answer #1”:



And after selecting “Answer #2”:



The child question with no parent answer will appear when either answer is selected. The child question with "Answer #2" as the parent answer only appears when "Answer #2" is selected.

You can also see that these questions were too long, and the user cannot read the whole question. Try to keep your question below 45 characters.

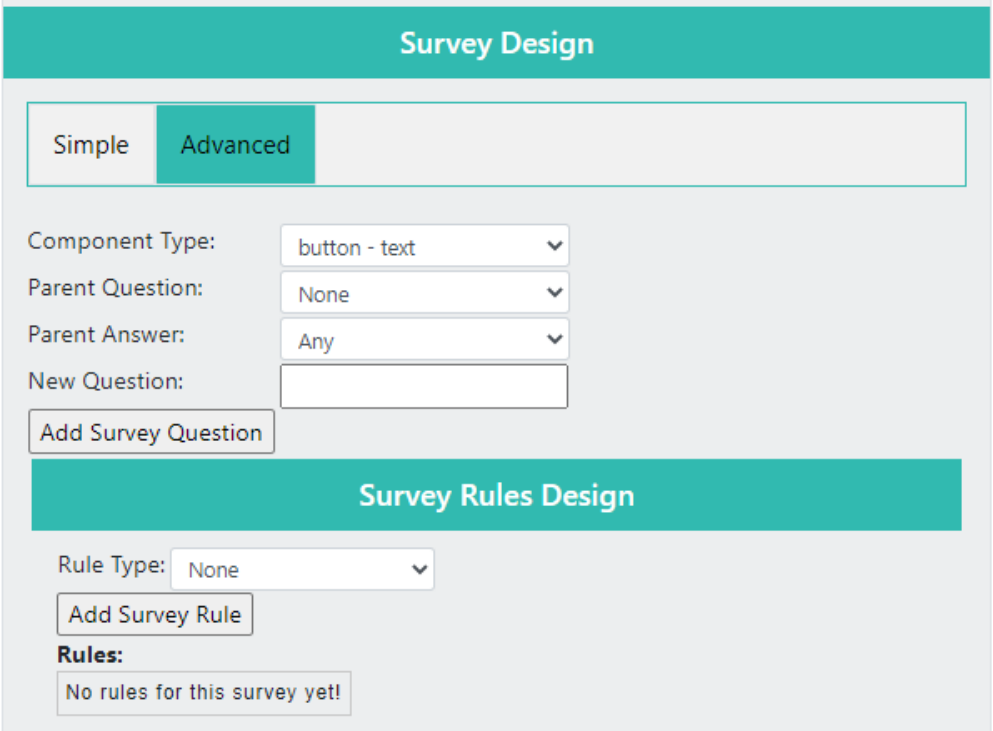
You can create extensive lists of land cover classification options for data collection using this approach.

If this type of question is enough for your project, proceed to Part 4: J Create Project after entering all your questions.

J. Survey design—advanced interface

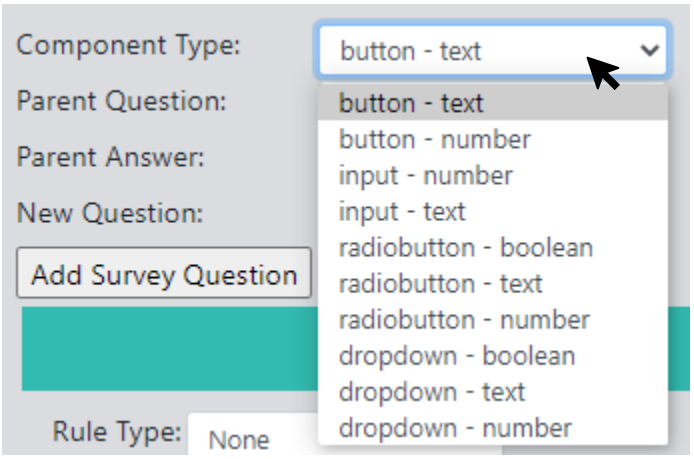
If you need more advanced functionality, use the **Advanced** interface under **Survey Design**.

With the **Advanced** interface, you can ask multiple types of questions (including the button—text questions from the **Simple** interface). You can also add survey rules in the **Survey Rules Design** panel.



Adding questions, assigning parent questions, assigning parent answers, and adding answers to questions all works as in the **Simple** interface. Deleting questions and collapsing Survey Cards is also the same. Please see Part 4: H Survey Design—Simple interface for more details.

The **Advanced** interface allows you to ask the following types of questions by clicking on the dropdown menu next to **Component Type:**



These **Component Types** include combinations of four question types and three data types.

The four question types are:

- **Button**: This creates clickable buttons, allowing users to select one out of many answers for each sample point.
- **Input:** Allows users to enter answers in the box provided. The answer text provided by the project creator becomes the default answer.
- **Radiobutton:** This creates radiobuttons, allowing users to select one out of many answers for each sample point.
- **Dropdown:** Allows users to select from a list of answers.

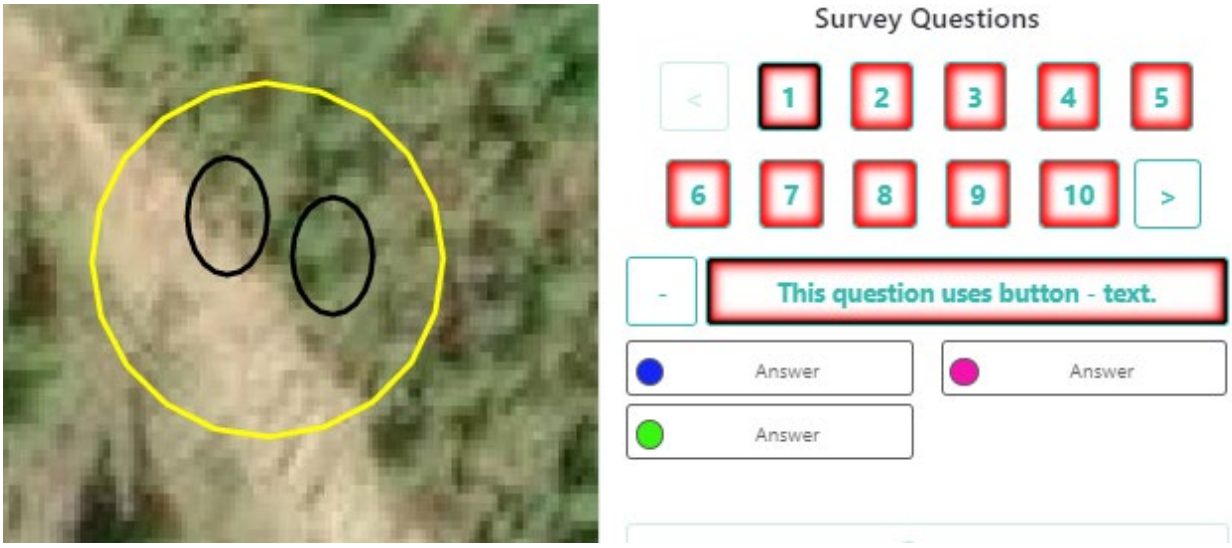
The three data types allowed are:

- **Boolean:** Use this when you have two options for a question (yes/no).
- **Text:** Use this when you have multiple options which are text strings. They may include letters, numbers, or symbols.
- **Number:** Use this when you have multiple options that are numbers, which do not contain letters or symbols.

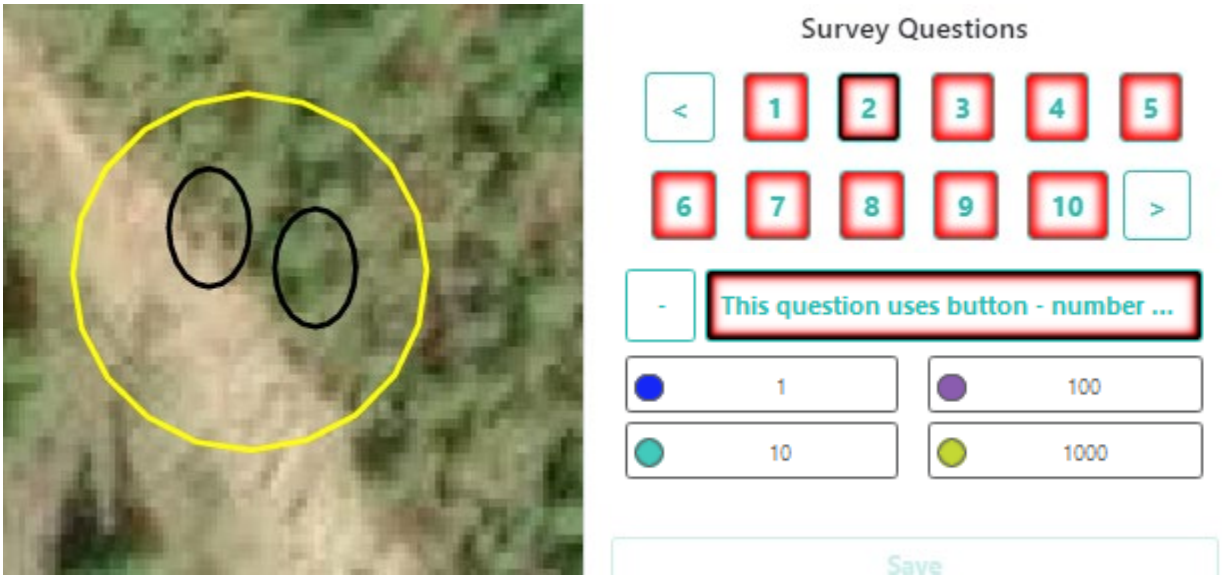
> Currently, you can have 'number' answers that include letters and symbols. If you include letters and symbols, you will not be able to implement number-based project rules for those questions.

Following are examples of how each question type listed under **Component** Type in the **Advanced** interface appears in the **Data Collection** screen, and notes on when each type might be useful.

Button—text:



Button—number:



Button—text and **Button—number** are useful when you want the user to classify each sample point as one of many different options. Different sample points can be assigned different answers. These are a great option for land use and land cover questions.

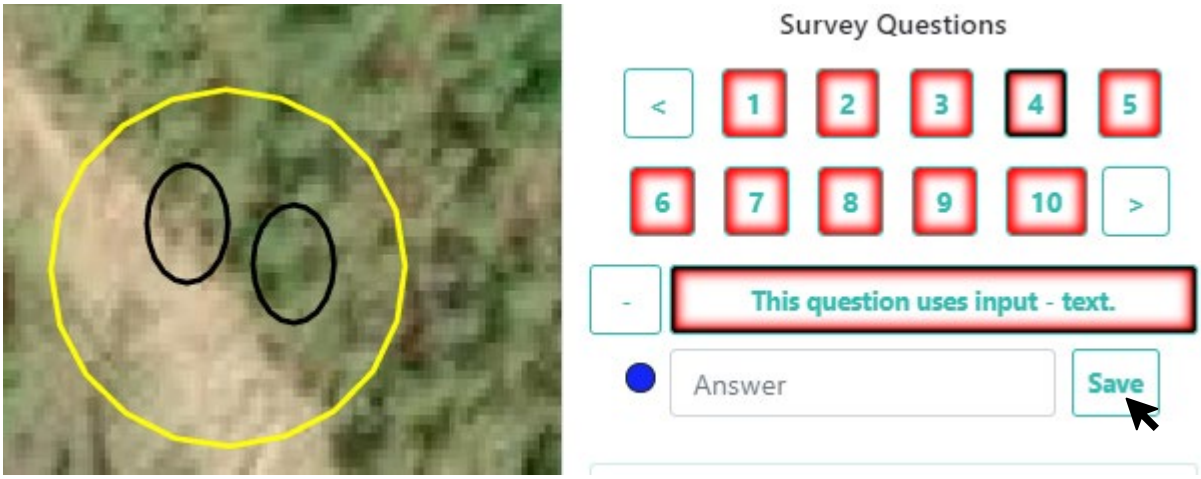
Try to use answer names with 15 characters or less so that the full name is displayed during analysis.

The **Button—number** option, coupled with numeric answers, allows you to implement numeric-only rules that will not work on **Button—text** elements.

Input—number:

Users can input integers, decimals, negative and positive numbers, and the letter “e” for scientific notation. Decimals must use “.” and not “,”. What the survey creator types in the “Answer” field will become the default text in the input box. Note the user must click “Save” to input the answer.

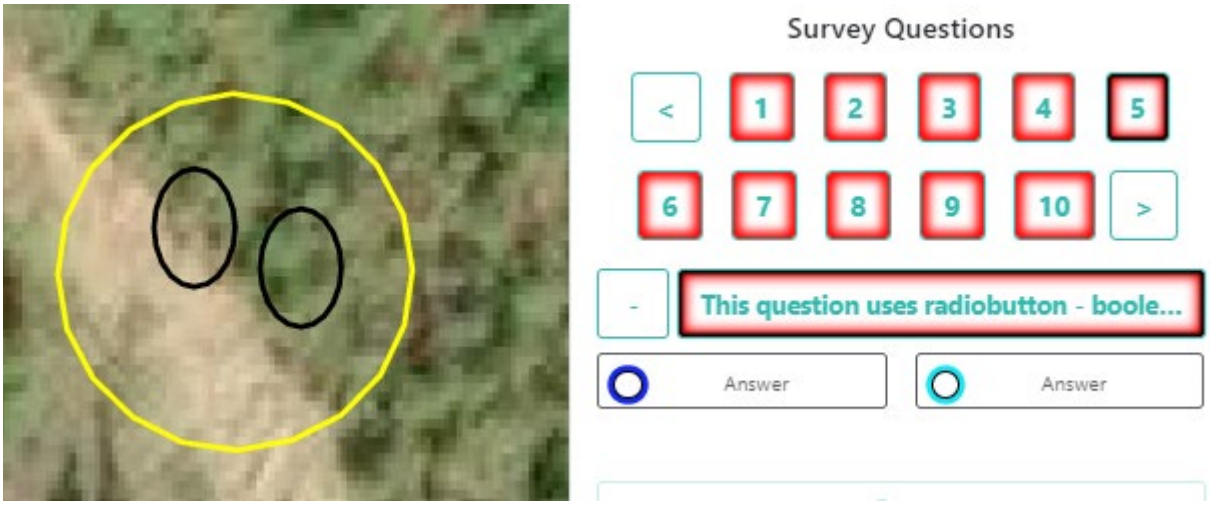
Input—text:



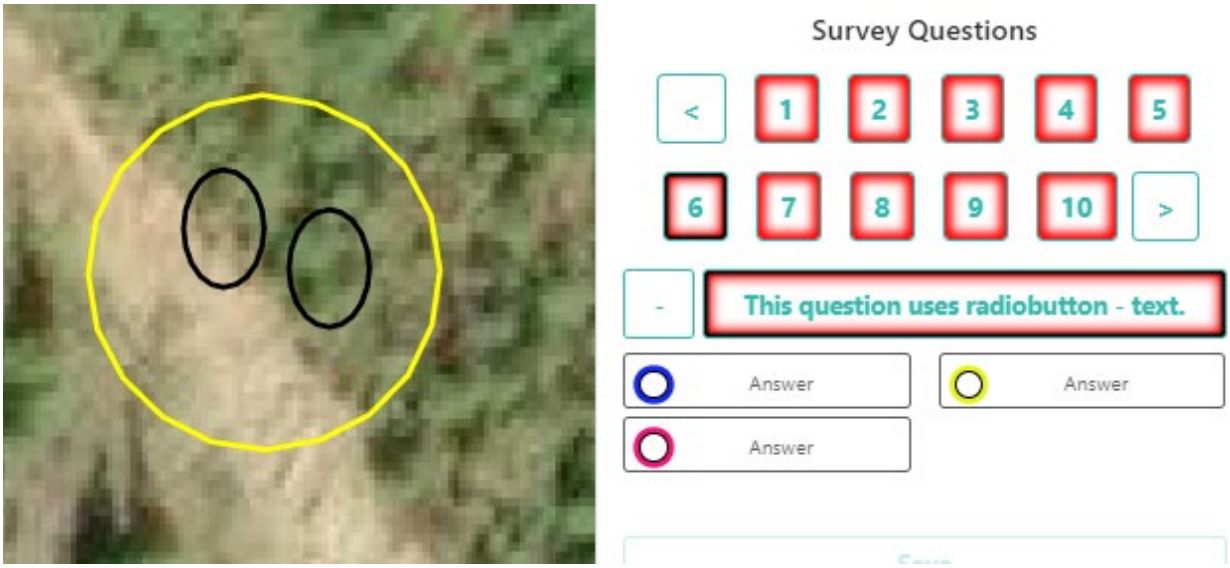
Users can input any character. What the survey creator types in the “Answer” field will become the default text in the input box.

Input—number and **Input—text** are useful when you want the user to provide custom input for each plot. For example, you might ask local participants to identify agricultural crops. Answers can be long (500+ characters).

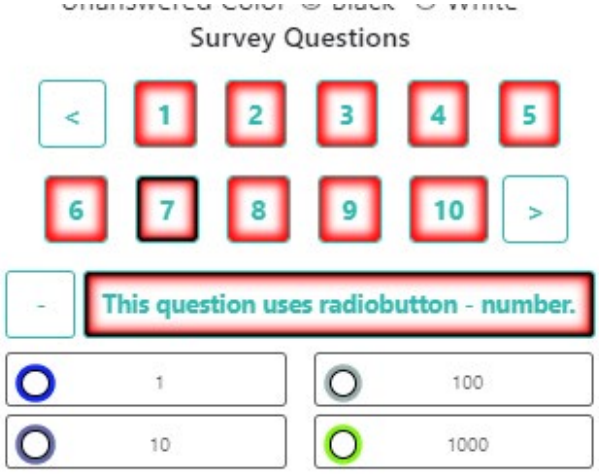
Radiobutton—boolean:



Radiobutton—text:



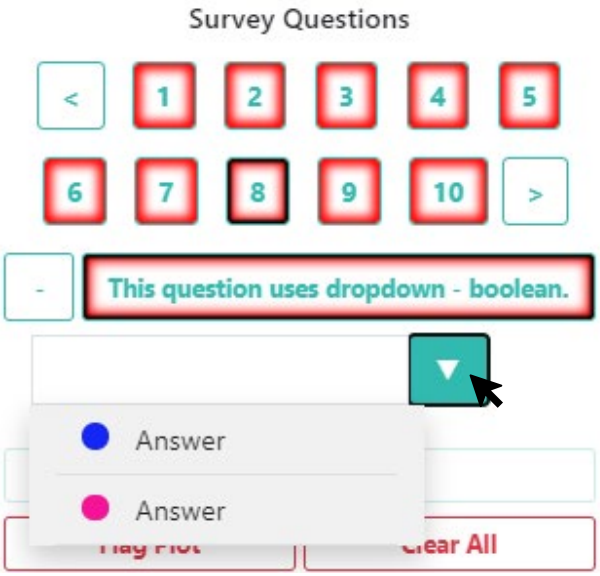
Radiobutton—number:



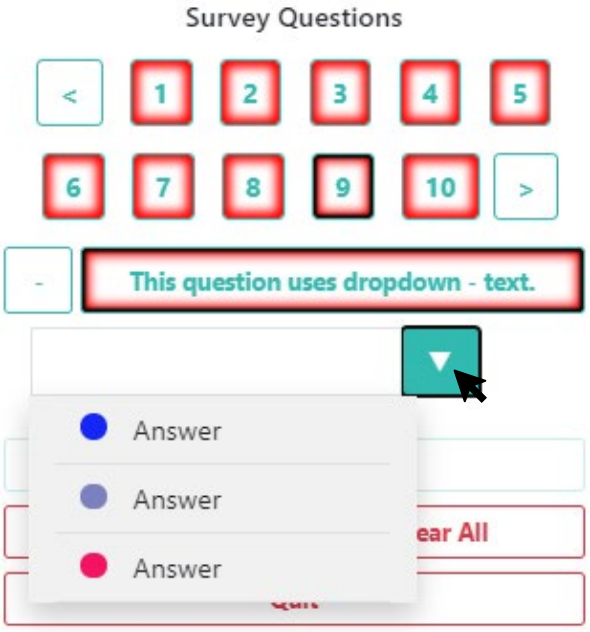
Radiobuttons are functionally much the same as Buttons, but with a different aesthetic and the added Boolean option. You can choose different answers for each sample point. Try to use names with around 15 characters or less so that the full name is displayed during analysis.

Radiobuttons are useful when you want your users to choose one answer for each of your sample points from multiple options. They work well for land use and land cover questions. The Boolean option also works well for areas you want to classify as yes/no, e.g. Forested or Not Forested.

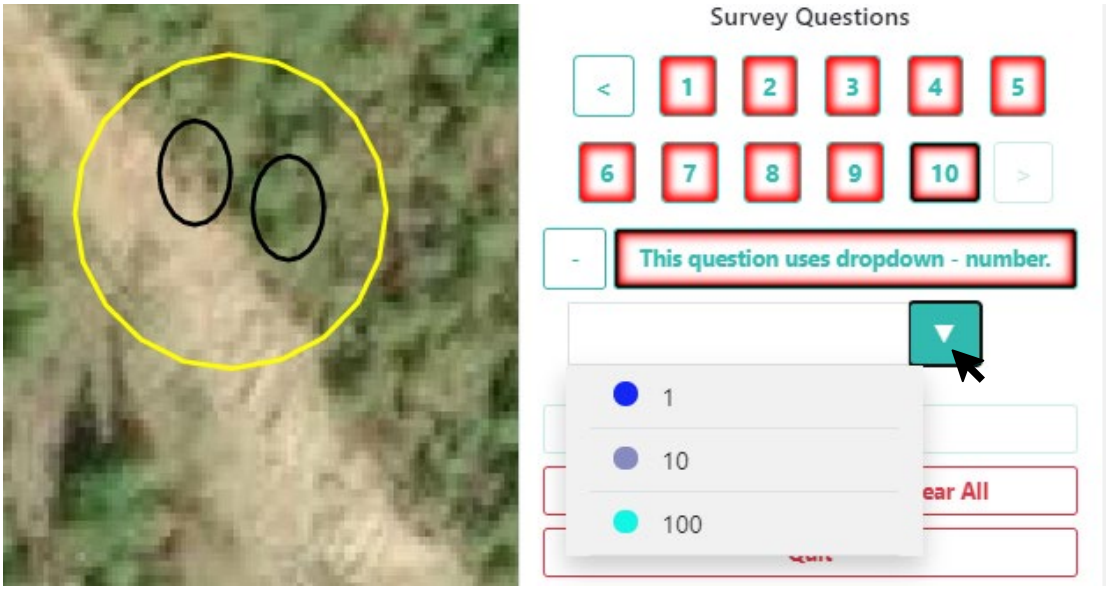
Dropdown—boolean:



Dropdown—text:

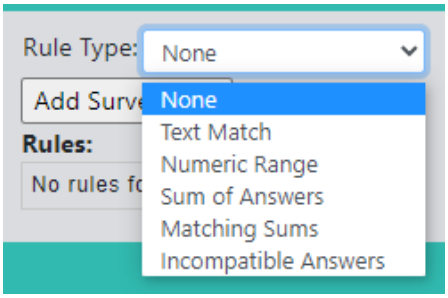


Dropdown—number:



The **Dropdown** questions function similarly to the **Button** and **Radiobutton** options, but with the dropdown menu aesthetic instead of the button aesthetic. As with the other options, you can assign each sample point a different answer, though it is more difficult from a user perspective. This option may encourage assigning only one answer to all the sample points. Overall, choose the option that will be easiest for your users to understand.

The **Advanced Interface** also allows you to specify survey rules.

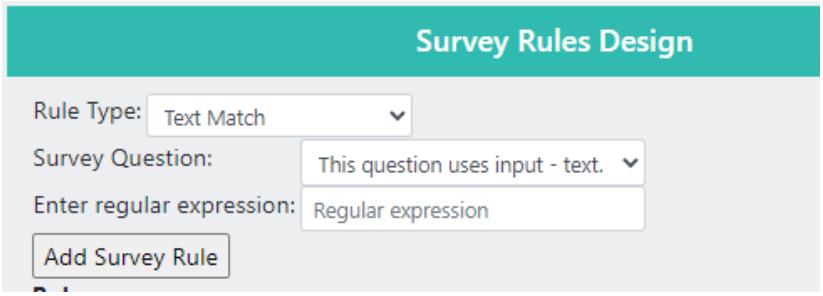


Rule Types include:

Text Match: This rule applies only to **Input—text** questions & their answers. It allows you to verify if the entered value fits, using regular expressions.

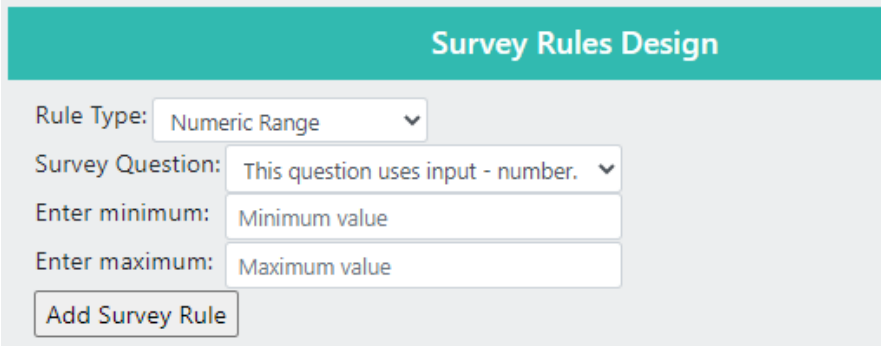
However, unless you have a specific reason to use the **Input—text** question type, consider using **Button—text** or **Radiobutton—text** instead. These options are easier for users and will always provide exact text.

This rule uses the JavaScript RegExp function, documentation for writing a regular expression can be found here: https://developer.mozilla.org/en-US/docs/Web/JavaScript/Guide/Regular_Expressions.



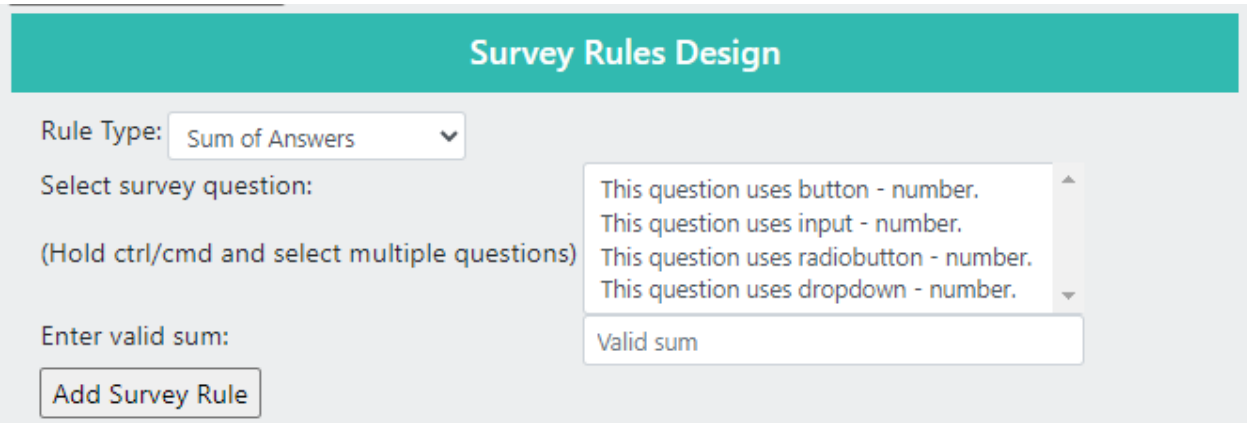
Numeric Range: This rule applies to **Input—number** questions & their answers. With this rule, you can verify that the numeric input falls within a predefined range.

For example, if you are asking about the proportion of points in the plot that contain trees, you could constrain the answers between 0 and 1.

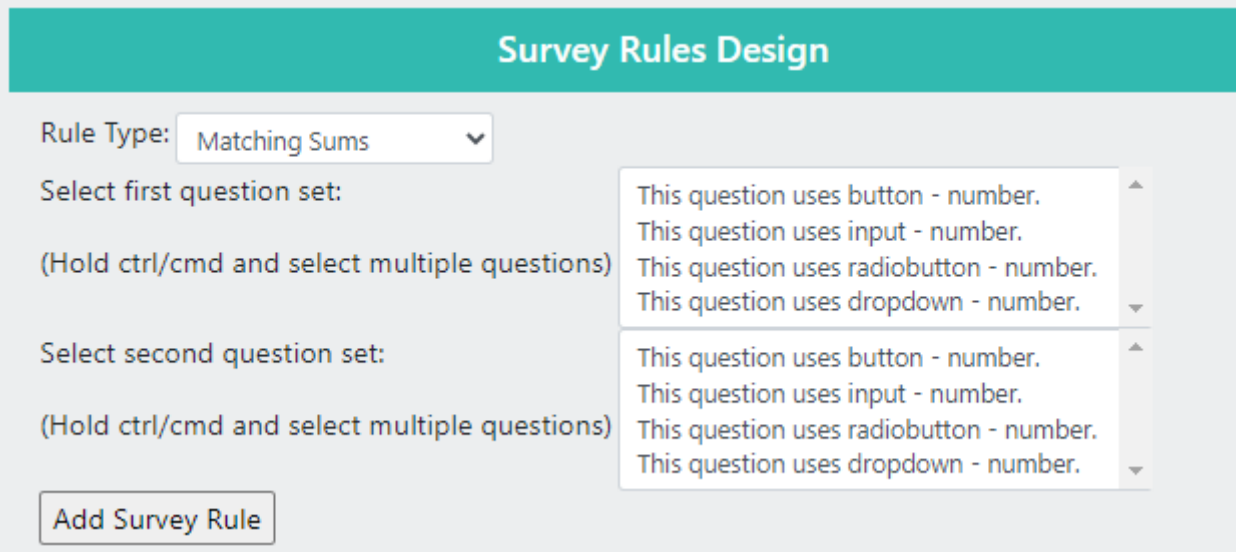


Sum of Answers: This rule applies to any **number** type questions & their answers. You select multiple questions (2 or more) and specify what the questions should sum to.

For example, this is helpful if you have multiple questions asking about percent of land cover, where the sum should be 100%.

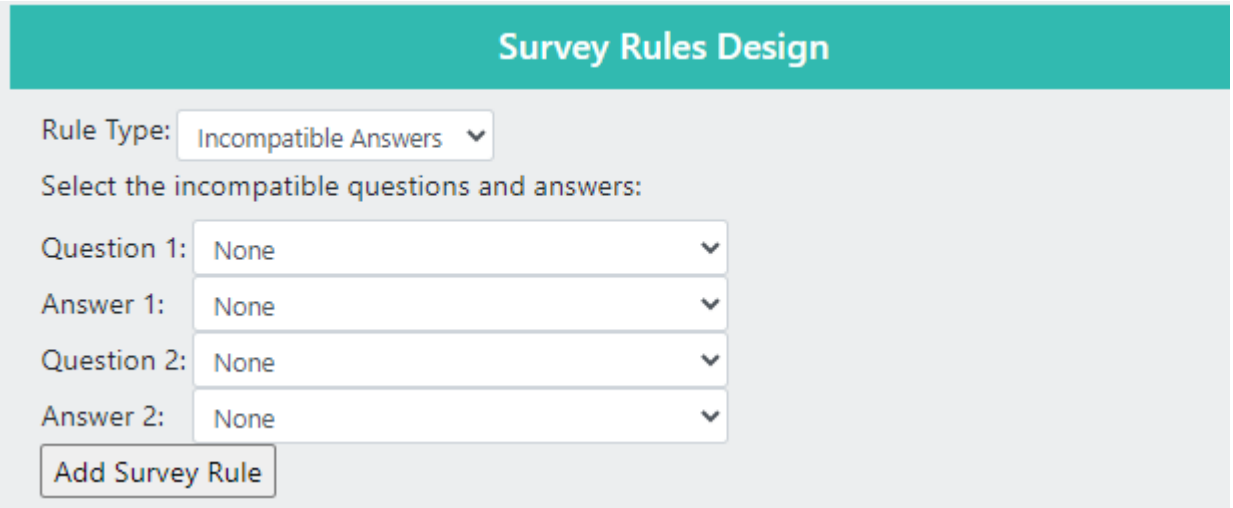


Matching sums: This rule applies to any **number** type questions & their answers. With this rule, you specify two sets of multiple questions (2 or more) that should have equal sums.

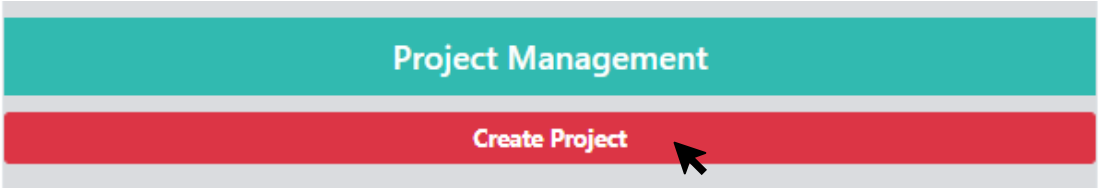


Incompatible answers: This rule can apply to any type of question. It allows the user to define incompatible sets of answers.

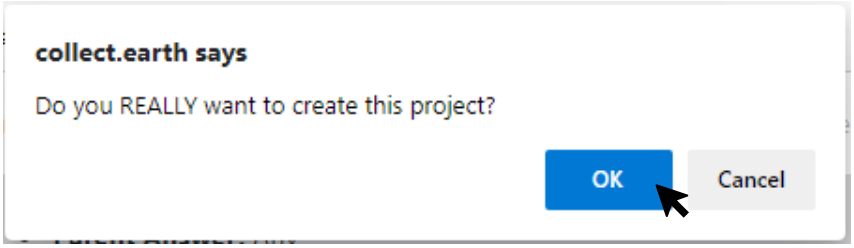
For example, if the answer to one question is land cover = “Water”, the answer to another question could not be land use = “Industrial”.



K. Create project



1. Once the project set up is complete, click on **[Create Project].**
2. A pop-up window will ask 'Do you REALLY want to create this project?' Click **OK**.



3. After a short delay, CEO will take you to the Review Project page (Part 5 following).

Part 5: Review Project

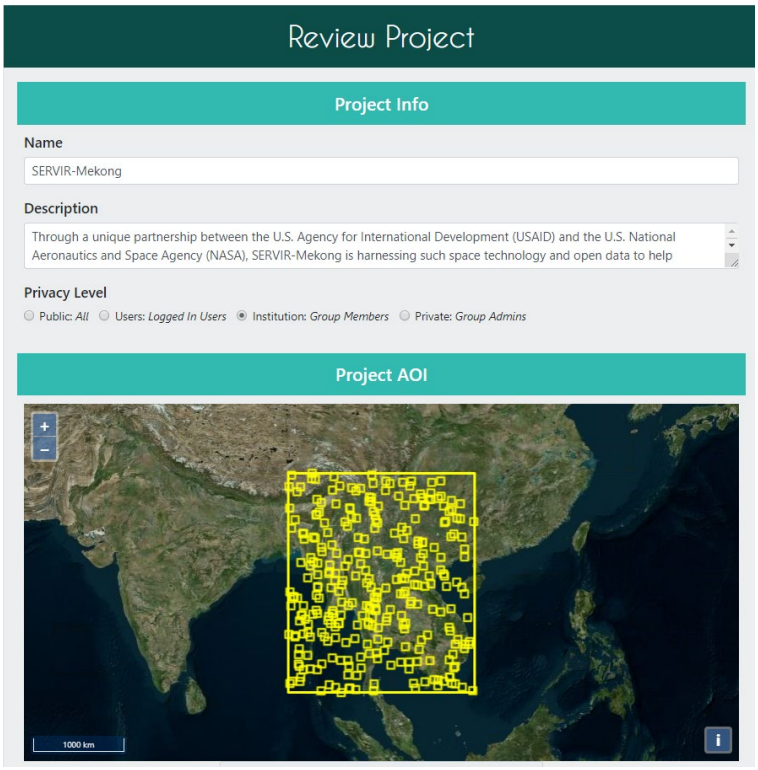
After you create the project, you will see the **Review Project** page (example on next page).

The only options you can change currently are under **Project Info**. They are the **Project Name, Description, and Privacy Level.** After you change any of these options you will need to click on **Update Project** at the bottom of the page.

Currently, project review is limited to the Project Info. If you see an error if your Project AOI, Plot Review, Sample Design, or Survey Review, you will need to start over by creating a new project.

If you publish your faulty project, you can use it as the template for a revised project and not lose all your work.

CEO is in the process of implementing project review functionality. Once implemented, this will allow Institution admins & reviewers (a new user class) to provide feedback on and alter the Survey Design and other aspects of the project prior to publishing the project.

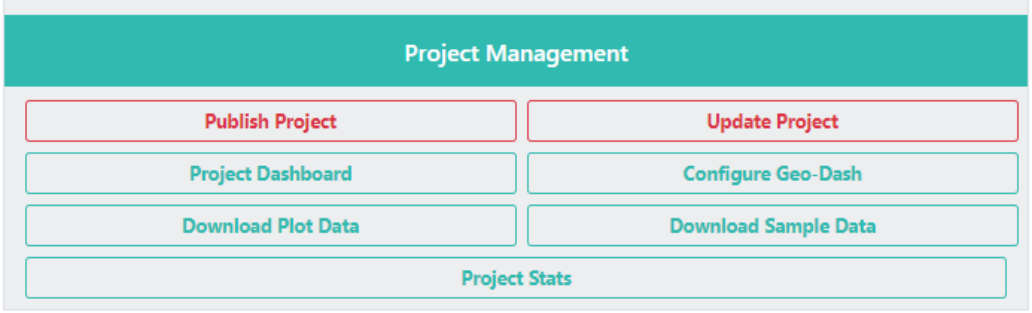


The **Project AOI** will now show the location of a subset of your plots (a maximum number can be displayed).

Not shown in the image above are the **Plot Review, Sample Design**, and **Survey Review** sections. The **Plot Review** and **Sample Design** sections will show a summary of the choices you made or the .csv and .shp files you uploaded. **Survey Review** shows all the **Survey Cards** you created, along with the corresponding **Component Type, Rules,** and **Answers**.

At this point, your project has been created, but it has not been published so that other users can see it. You can either click **[Publish Project]** or **[Configure Geo-Dash]**. The option to **Configure Geo-Dash** will be available after you publish your project, as well. The following sections discuss Geo-Dash (Part 6) and published project management (Part 7).

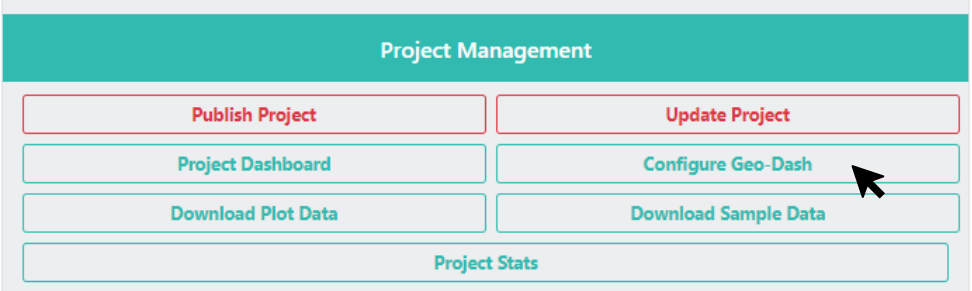
Note that if a project is not "Published" ONLY THE ADMINS OF THE INSTITUTION CAN SEE IT, NOT THE MEMBERS.



Part 6: Geo-Dash Implementation

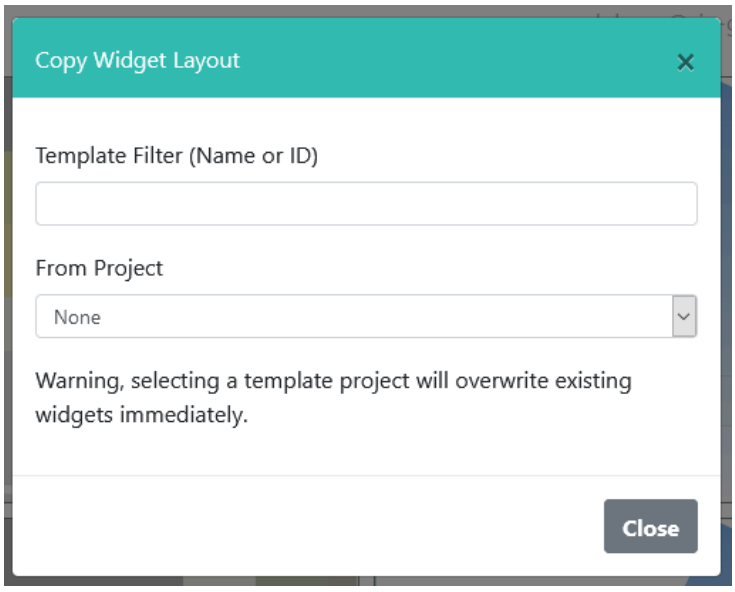
Geo-Dash is a dashboard that opens in a second window when users begin to analyze sample plots. Geo-Dash provides users with additional information to help them interpret the imagery and better classify sample points and plots. The Geo-Dash tab can be customized to show information such as NDVI time series, additional imagery, and digital elevation data.

You can configure your **Geo-Dash** by clicking on **[Configure Geo-Dash]** from the **Project Information/ Review Project** page**.**



This will bring up the **Geo-Dash layout screen**.

[Copy Layout] will allow you to copy the Geo-Dash from another project. This will delete any Geo-Dash you have associated with the project!



You can add individual Geo-Dash widgets with the **[Add Widget].** To find out what each of the widgets do, clicking on **[Geo-Dash Help]** will open the **Geo-Dash Help Center**. We will also discuss this functionality below.

NDVI, EVI, EVI 2, NDMI, and NDWI indices are available in both image overlays and time series graphs (sections A, B, D below). We provide some basic information about these indices here; more information is readily available.

The normalized difference vegetation index (NDVI) is used to determine if the cell contains live green vegetation. In map based representations, dead plants and inanimate objects are represented as red, while live healthy plants are represented as green. In numerical representations (e.g. time series graphs), values below 0 represent dead plants or inanimate objects, 0-0.33 represents unhealthy plants, 0.33-0.66 represents moderately healthy plants, and 0.66-1 represents very healthy plants. For more information, see e.g. https://www.usgs.gov/land-resources/nli/landsat/landsat-normalized-difference-vegetation-index

The enhanced vegetation index (EVI) and two-band EVI (EVI 2) are optimized vegetation indexs. They are designed to have higher sensitivity in high biomass regions e.g. along the equator, correct for canopy background signals, and reduce atmospheric influence on index values. In doing so, these indices addresses some of the key limitations of NDVI, however EVI requires more data to calculate and therefore has its own limitations. EVI 2 is in development and can be calculated just from red and near infrared bands. As with NDVI, red is used to represent dead plants/inanimate objects and green to represent healthy plants. The index varies between 0-1, with 0 representing dead plants and 1 representing very healthy plants. For more information on EVI see e.g. https://www.usgs.gov/land-resources/nli/landsat/landsat-enhanced-vegetation-index.

The normalized difference moisture index (NDMI) is used to determine the water content of vegetation. NDMI can be used for drought monitoring and for determining fuel loads (combustability) for wildfire hazard assessments. Values near -1 indicate plants with low moisture while values near 1 indicate plants with high moisture. More information on NDMI can be found at e.g. https://www.usgs.gov/land-resources/nli/landsat/normalized-difference-moisture-index.

The normalized difference water index (NDWI) is also related to plant water content and plant water stress. It can be used to map water bodies, determine crop health and for wildfire risk analysis. Values near -1 indicate low water content and vegetation cover and values near 1 indicate high water content and vegetation cover. More information on NDWI can be found at e.g. https://edo.jrc.ec.europa.eu/documents/factsheets/factsheet_ndwi.pdf.

Data from Landsat satellites 5, 7, 8 and Sentinel 2 are also available. For more information about these datasets, see: Landsat: https://www.usgs.gov/land-resources/nli/landsat and Sentinel 2: https://sentinel.esa.int/web/sentinel/missions/sentinel-2.

CEO uses Google Earth Engine to process many of the Geo-Dash widgets. Therefore, the information about Image Visualization for GEE is also useful here, specifically the min and max descriptions: https://developers.google.com/earth-engine/image_visualization

A. Widget formatting notes

1. **Titles** cannot contain special characters.
2. The Custom Widgets are useful but require some specific syntax. Quotes used for **custom widgets** should be vertical (not inverted comma style). There should be no spaces in the **custom widget Image parameters**. If you do not have time to export mosaics to GEE Assets, you can create mosaics on the fly in the GEODASH:
3. For the **Date Ranges**, if the end date is longer than the period of record, only the available data will be displayed.

B. Add an Image Collection Widget

Image Collection Widgets provide additional information about the survey plots and points.

1. Click **[Add Widget]** in the upper right hand of the **Geo-Dash layout screen.**
2. Select **Image Collection** in the **Type** dropdown menu.
3. Choose your **Basemap** source from the dropdown menu. This will become the basemap over which the other data is overlaid.
4. Select **Data**:
    i. **NDVI, EVI, EVI 2, NDMI**, and **NDWI** are preconfigured with the correct Landsat bands and image parameters. NDVI merges l4 thru l8 and sentinel 2.

(a) If you select one of these, you will just need to add a **Title** for the widget. We suggest an informative title such as {Data} {Date range} e.g. NDVI 2001 or NDVI 2001 - 2002

(b) You will also need to add the **Date Range**. The index displayed will be the mean for the selected date range.

(c) You can overlay an additional date range for comparison. If you would like to enable this feature, tick the **Dual time span** checkbox, and select a second date range.

ii. **Landsat 5, Landsat 7, Landsat 8, and Sentinel 2** are partially configured leaving you the option to adjust the bands, min, max, and cloud score. Available bands for each are:

(a) Landsat 5 - B1, B2, B3, B4, B5, B6, B7, BQA

(b) Landsat 7 - B1, B2, B3, B4, B5, B6_VCID_1, B6_VCID_2, B7, B8, BQA

(c) Landsat 8 - B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, BQA

(d) Sentinel 2 - B1, B2, B3, B4, B5, B6, B7, B8, B8a, B9, B10, B11, B12, QA10, QA20, QA60

(e) Min, Max, and Cloudscore will be determined by the user's needs. Min and Max are the values for the bands, while cloud score is a max percent of clouds in the image. You must have one value of Min and Max for each band. Example values include:

(i) Landsat5:
bands: "B4,B5,B3",
min: "0.05,0.01,0.07",
max: "0.45,0.5,0.4",
cloudLessThan: 90

(ii) Landsat7:
bands: "B4,B5,B3",
min: "0.03,0.01,0.05",
max: "0.45,0.5,0.4",
cloudLessThan: 90

(iii) Landsat8:
bands: "B5,B6,B4",
min: "0.03,0.01,0.04",
max: "0.45,0.5,0.32",
cloudLessThan: 90

(iv) Sentinel:
bands: "B8,B4,B3",
min: "900,450,800",
max: "5200,3000,2000",
cloudLessThan: 10

(f) The pieces that are preconfigured include:

(i) landsat8: imageCollection: LANDSAT/LC08/C01/T1_RT; simpleCompositVariable: 50

(ii) landsat7: imageCollection: LANDSAT/LE07/C01/T1; simpleCompositVariable: 60

(iii) landsat5: imageCollection: LANDSAT/LT05/C01/T1; simpleCompositVariable: 50

(iv) sentinel: imageCollection: COPERNICUS/S2

iii. **Custom widget** - Any collection from Google Earth Engine can be added if you know the dataset. You simply need to know the image name and the image parameters you would like. For example:

(a) Offline Carbon Monoxide - COPERNICUS/S5P/OFFL/L3_CO {"bands":"CO_column_number_density,H2O_column_number_density,cloud_height","min":"0","max":"0.5"} 2018-01-01 to 2018-12-31

(b) Notice the double quotes in the image parameters for both the property name and the value. Also, notice there are no spaces.

(c) The vision parameters are similar but slightly different than you would add directly in the Google Earth Engine code editor. For example, the bands parameter in the editor is an array of comma-separated strings, and here it is a single comma separated string.

(d) For better performance when using a GEE collection, we recommend that you preprocess and clip the data to your AOI in GEE before connecting it to Geo-Dash. You can find a video tutorial on this process in the Geo-Dash Help Center under "Video Tutorial to Create GEE Image Assets."

5. Select **Date Range.** If the end date is longer than the period of record, only the available data will be displayed.
6. Click Create.
7. Reposition and resize to your liking.
8. For a .gif illustrating this process, see: https://collect.earth/img/image_collection_widget.gif

C. Add a Time Series Graph Widget

Time Series Graphs, particularly for the vegetation indices, can help users differentiate between different types of vegetation based on seasonal patterns.

1. Click **[Add Widget]** in the upper right hand of the **Geo-Dash layout screen.**
2. Select **Time Series Graph** in the **Type** drop down.
3. Select **Data**:
   i. **NDVI, EVI, EVI 2, NDMI**, and **NDWI** are preconfigured with the correct Landsat band calculations and image parameters. NDVI merges Landsat-4 thru Landsat 8 and Sentinel 2.
      (a) If you select one of these, you will just need to add a **Title** for the widget. We suggest an informative title such as {Data} {Date range} e.g. NDVI 2001 or NDVI 2001 - 2002
      (b) You will also need to add the **Date Range**.
   ii. **Custom widget** - Any collection from Google Earth Engine can be added if you know the dataset. You simply need to know the image name, the band you would like graphed, and how you would like the graph reduced. For example:
      (a) **GEE** Image Collection - COPERNICUS/S5P/OFFL/L3_CO.

(b) Band to graph - CO_column_number_density.
(c) Select the **Reducer** in the dropdown.

4. Select a **Date Range**. If the end date is longer than the period of record, only the available data will be displayed.
5. Click **Create**.
6. Reposition and resize to your liking.
7. For a gif illustrating this, see: https://collect.earth/img/time_series_graph_widget.gif

D. Add a Statistics Widget

Statistics provide additional information, including population, area, and elevation about the current plot area that is being analyzed. The population data is from the CiESIB Gridded Population of the World v4 dataset and the elevation data is from the USGS EROS Global 30-Arc Second DEM.

1. Click **[Add Widget]** in the upper right hand of the **Geo-Dash layout screen.**
2. Select **Statistics** in the type drop down.
3. Give the widget a **Title.**
4. Click Create.
5. Reposition and resize to your liking.
6. For a gif illustrating this, see https://collect.earth/img/statistics_widget.gif

E. Add a Dual Image Collection Widget

The dual image collection widget combines two different image collection widgets so the user can compare the images side by side. Refer to Part 6: B “Add an Image Collection Widget” for more detailed information.

1. Click **[Add Widget]** in the upper right hand of the **Geo-Dash layout screen.**
2. Select **Dual Image Collection** in the **Type** drop down.
3. Choose **Basemap** source from dropdown. This will be the basemap for the widget and other data will be layered on top.
4. Under Dual imageCollection Step 1, select Data:
   i. **NDVI, EVI, EVI 2, NDMI**, and **NDWI** are preconfigured with the correct Landsat bands and image parameters. NDVI merges l4 thru l8 and sentinel 2.
      (a) If you select one of these, you will just need to add a **Title** for the widget. We suggest an informative title such as {Data} {Date range} e.g. NDVI 2001 or NDVI 2001 - 2002
      (b) You will also need to add the **Date Range**.

The Geo-Dash interface will give you an option to overlay an additional date range for comparison via the **Dual time span** checkbox.

This function does not work.

ii. **Landsat 5, Landsat 7, Landsat 8**, and **Sentinel 2** are partially configured leaving you the option to adjust the bands, min, max, and cloud score. Available bands for each are:
    (a) Landsat 5 - B1, B2, B3, B4, B5, B6, B7, BQA
    (b) Landsat 7 - B1, B2, B3, B4, B5, B6_VCID_1, B6_VCID_2, B7, B8, BQA
    (c) Landsat 8 - B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, B11, BQA
    (d) Sentinel 2 - B1, B2, B3, B4, B5, B6, B7, B8, B8a, B9, B10, B11, B12, QA10, QA20, QA60
    (e) For information on the min, max, and cloud score see Part 6: B "Add an Image Collection Widget". Min and Max are the values for the bands, while cloud score is a max percent of clouds in the image.

iii. **Image Asset** (see Image Asset example in Part 6: E "Add an Image Asset Widget" and skip step 5 below as you will not need to select a date range)

iv. **Image Collection Asset** (see Image Collection example in Part 6: F "Add an Image Collection Asset.)

v. **Custom widget** - Any collection from Google Earth Engine can be added if you know the dataset. You simply need to know the image name and the image parameters you would like for example:
    (a) Offline Carbon Monoxide - COPERNICUS/S5P/OFFL/L3_CO
    {"bands":"CO_column_number_density,H2O_column_number_density,cloud_height","min":"0","max":"0.5"} 2018-01-01 to 2018-12-31
    (b) Notice the double quotes in the image parameters for both the property name and the value. Also, notice there are no spaces. The vision parameters are similar but slightly different than you would add directly in the Google Earth Engine code editor. For example, the bands parameter in the editor is an array of comma-separated strings, and here it is a single comma separated string

5. Select **Date** range, if applicable. If the end date is longer than the period of record, only the available data will be displayed.

6. Click **[Step 2 =>].** You can navigate back to Step 1 by clicking **[<= Step 1]** after going to Step 2.

7. Select **Data 2**.
    i. The options are the same as for **Data 1**, above.
    ii. However, the option to add a second **Date Range** (not currently functional) is not available.

8. Select **Date** range for **Data 2.**

9. Click Create.

10. Reposition and resize to your liking.

11. For a gif illustrating this widget, see: https://collect.earth/img/dual_image_collection_widget.gif

F. Add an Image Asset Widget

Both the image asset widget and the image collection asset widget allow users to connect to GEE data. This includes publicly available data through GEE as well as image assets from their own user accounts. There is a video tutorial for creating GEE image assets available online here: https://youtu.be/l57IhmduVBQ. Some example GEE scripts to create Image Asset in GEE:

- NDVI: https://github.com/SERVIR/gee-scripts/blob/master/GEE_NDVI_Clip_to_plot_Collection.js#L6
- Landsat8: https://code.earthengine.google.com/109485ad0712818dba270b767efbc540

1. Click **[Add Widget]** in the upper right hand of the **Geo-Dash layout screen.**
2. Select **Image Asset** in the **Type** drop down.
3. Choose **Basemap** source from dropdown. This will be the basemap for the widget and other data will be layered on top.
4. Give the widget a **Title**.
5. Enter the **GEE Image Asset** - Example: projects/servir-e-sa/…
6. Enter **Image Parameters** for asset. – For example:
   i. {"bands":"CO_column_number_density,H2O_column_number_density,cloud_height","min":"0","max":"0.5"}
   ii. Note that there are no spaces.
7. You can also add the **Asset to your institution basemaps** using the checkbox.
8. Click Create.
9. Reposition and resize to your liking.
10. For a gif illustrating this widget, see: https://collect.earth/img/image_asset_widget.gif.

More on Landsat: The most common strings for Landsat are listed below, detailed information can be found on Google Earth Engine help pages. Available composites and the covered time frame are found here: https://explorer.earthengine.google.com/#search/tag%3A32day.

For short change intervals, test the **8 day** NDVI composite images; for longer change intervals, explore the **32 day** composites.

You might notice in the 8 day composites unexpectedly low NDVI values for the season. This can be caused by **cloud cover**. If your sample sites are in an area with persistent cloud cover, choose the 32 day composite.

Satellite	Type	Start date	Image collection ID
Landsat 8	NDVI, 32 day	2013-04-07	LANDSAT/LC8_L1T_32DAY_NDVI
Landsat 8	NDVI, 8 day	2013-04-07	LANDSAT/LC8_L1T_8DAY_NDVI
Landsat 7	NDVI, 32 day	1999-01-01	LANDSAT/LE7_L1T_32DAY_NDVI
Landsat 7	NDVI, 8 day	1999-01-01	LANDSAT/LE7_L1T_8DAY_NDVI
Landsat 8	NDWI, 32 day	2013-04-07	LANDSAT/LC8_L1T_32DAY_NDWI
Landsat 8	NDWI, 8 day	2013-04-07	LANDSAT/LC8_L1T_8DAY_NDWI
Landsat 7	NDWI, 32 day	1999-01-01	LANDSAT/LE7_L1T_32DAY_NDWI
Landsat 7	NDWI, 8 day	1999-01-01	LANDSAT/LE7_L1T_8DAY_NDWI

The image collection ID for Sentinel-2 is COPERNICUS/S2 (2015-06-23 to present) and for Sentinel-1 is COPERNICUS/S1_GRD (2015-10-03 to present).

G. Add an Image Collection Asset Widget

An Image Collection is a stack or time series of images. There is a video tutorial for creating GEE ImageCollection assets available online here: https://youtu.be/7eIvltgDbXw.

1. Click **[Add Widget]** in the upper right hand of the **Geo-Dash layout screen.**
2. Select **Image Collection Asset** in the **Type** drop down.
3. Choose **Basemap** source from dropdown. This will be the basemap for the widget and other data will be layered on top.
4. Give the widget a **Title**.
5. Enter the **GEE Image Asset** - Example: users/ValeriaContessa/Indonesia_2000
6. Enter **Image Parameters** for asset in JSON format - Example:
   i. {"bands":"B4,B5,B3","min":"10,0,10","max":"120,90,70"}
   ii. Again, do not use any spaces.

7. You can also add the **Asset to your institution basemaps** using the checkbox.
8. Click Create.
9. Reposition and resize to your liking.
10. For a gif illustrating this widget, see https://collect.earth/img/image_collection_asset_widget.gif

H. Add SRTM Digital Elevation Data 30m Widget

The SRTM Digital Elevation Data 30m widget pulls data from the Earth Engine Data Catalog. More information on this dataset can be found here: https://developers.google.com/earth-engine/datasets/catalog/USGS_SRTMGL1_003. The SRTM is an image asset USGS/SRTMGL1_003. If you would like to have it more closely related to your AOI, you could clip the asset and save it to your GEE asset directory and then add it thru the image assets widget type.

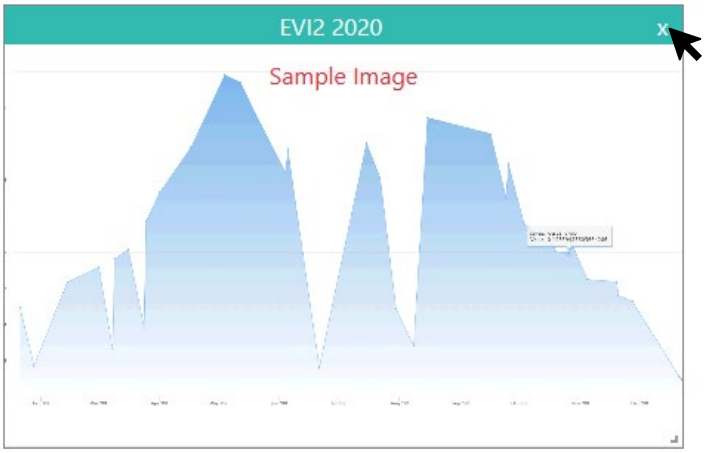
1. Click **[Add Widget]** in the upper right hand of the **Geo-Dash layout screen.**
2. Select SRTM Digital Elevation Data 30m in the Type drop down.
3. Choose the **Basemap** source from dropdown. This will be the basemap for the widget and other data will be layered on top.
4. Give the widget a **Title**.
5. Enter **Image Parameters** in JSON format.
    i. For example, {"min":"0","max":"600","palette":"blue,green,yellow,red"}
6. Click Create.
7. Reposition and resize to your liking.

I. Move & resize widgets

1. Widgets can be manipulated on the Geo-Dash Widget Layout Editor in multiple ways:
    i. Drag and drop
    ii. Resize by dragging from the bottom right corner
    iii. Widgets will auto align
2. Widgets are updated in real time
3. View rendered results in the Geo-Dash window
4. For a gif illustrating these movements, see https://collect.earth/geo-dash/geo-dash-help

J. Edit & delete widgets

1. You cannot edit widgets once they are created.
2. To delete a widget, click on the X in the upper right-hand corner.



K. Synthetic Aperture Radar (SAR) data in CEO

Observations of backscatter variations over time in satellite SAR data can be attributed to structure and moisture. For forests, these can be linked co changes in the moisture conditions of the trees and soil as well as changes in forest structure. These are extremely useful for e.g. detecting forest degradation.

Color display of this data is possible in CEO through GEE & Geo-Dash.

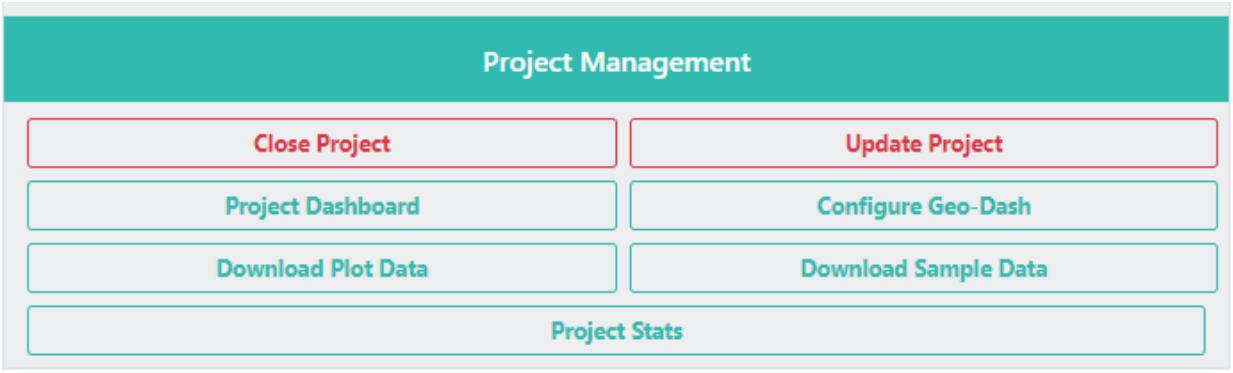
For more information, please see:

1. Kellndorfer, Josef. “Using SAR Data for Mapping Deforestation and Forest Degradation.” SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation. Eds. Flores, A., Herndon, K., Thapa, R., Cherrington, E. NASA. 2019. DOI: . 10.25966/68c9-gw82; available online at: https://gis1.servirglobal.net/TrainingMaterials/SAR/Ch3-Content.pdf
2. This one pager from SERVIR & SilvaCarbon: https://servirglobal.net/Portals/0/Documents/Articles/2019_SAR_Handbook/SAR_VegIndices_1page_new.pdf

Additional Geo-Dash widget functionality for detecting forest degradation via SAR.

Part 7: Published Project Management

The following options for project management are available on the **Project Information/Review Project** page after you have published your project:

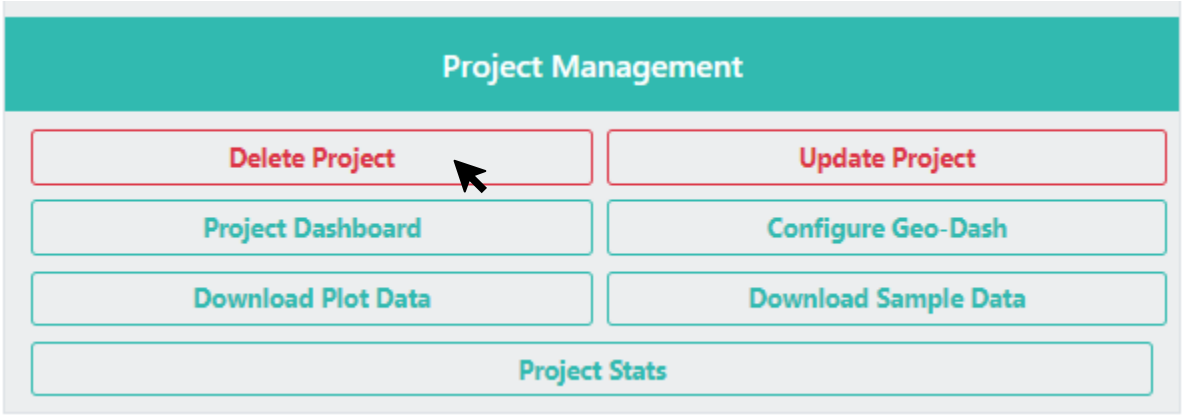


This menu is located at the bottom of the page. If you are navigating back to the **Project Information/Review Project** (i.e. not visiting right after project creation), you can reach this page by finding your Institution on the **Home** page left panel. Click the **[Edit]** icon ( ) next to the project to get to the **Project Information/Review Project** page.

A. Close Project & Delete Project

Clicking **Close Project** will close your project and prevent further data collection. You cannot reopen your project.

Once you close your project, you have the option to permanently delete your project by clicking **Delete Project** (see below).

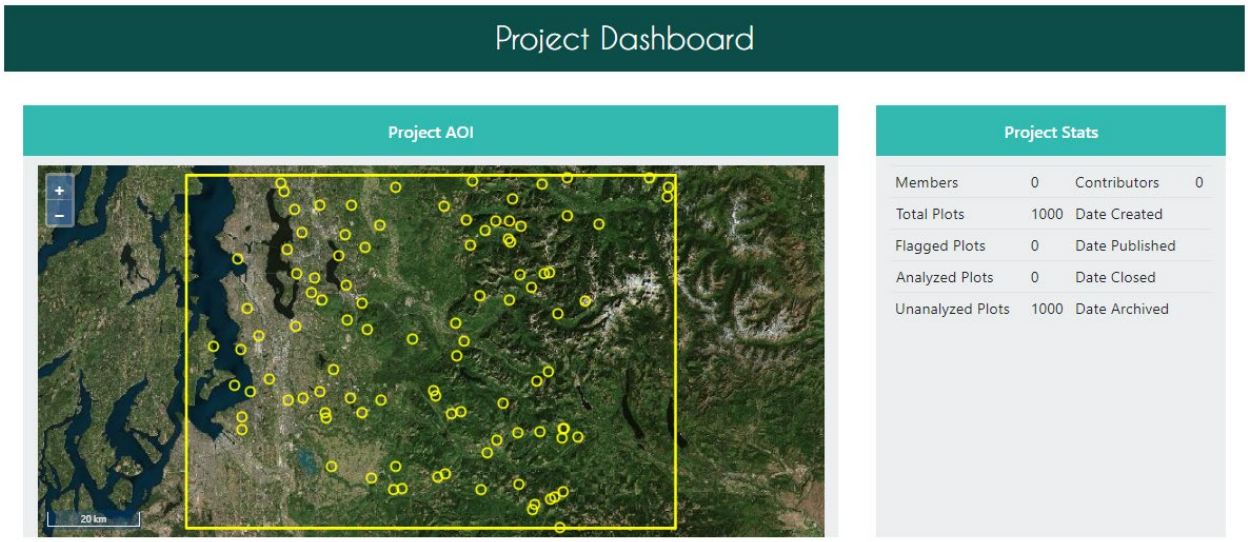


B. Update Project

The only options you can change currently are under **Project Info**. They are the **Project Name**, **Description**, and **Privacy Level**. After you change any of these options you will need to click on **Update Project** at the bottom of the page.

C. Project Dashboard

The **Project Dashboard** contains basic information about the project.



D. Configure Geo-Dash

Please see **Part 6: Geo-Dash Implementation** above.

E. Download your data

There are two data download options: **Download Plot Data,** which downloads your data summarized by plot, and **Download Sample Data**, which downloads your raw data, with information for each point within each plot as its own row. Both are downloaded in .csv, which can be opened in programs like Microsoft Excel or imported into data analysis software. Data downloaded from CEO will be in WGS84 EPSG:4326 format.

Downloaded .csv data from **Download Plot Data** will have the following columns:

1. **PLOT_ID**: the CEO-assigned unique sample plot number.
2. **CENTER_LON** and **CENTER_LAT** are the geographic coordinates of the center of your sample plots.
3. **SIZE_M** and **SHAPE** describe the size in meters and the shape (circle or square) of the sample plot.

4. **FLAGGED**: This will be FALSE for plots where data was collected and for plots where data has not been collected yet. It will be TRUE when a user has flagged the quality of the background map as not good enough to analyze the samples (e.g. due to clouds or poor image resolution).
    i. Plots can either be flagged or saved by a user, but not both.
    ii. If a user flags the plot, then goes back and enters data and saves the plot, the plot will not be flagged.
5. **ANALYSES** will show 1 for plots that are analyzed, and 0 stands for not analyzed. If more than one person assigned a label to a plot, this column will indicate the number of analysts that have assessed the plot.
6. **SAMPLE_POINTS** indicates the number of samples in each plot.
7. **USER_ID** is the user id (email address) of the person that classified the plot.
8. If you used a .csv or .shp file for plot design, any additional data columns you uploaded will be preserved in the .csv download. They will be preceded by **PL_(column name)**.
9. All the following columns will have information about each of the survey questions broken down by answer. They are labeled **QUESTION TEXT:ANSWER TEXT**. In the plot summary download, these are quantified as percent (max 100) of the sample points in the plot that were assigned that answer. For example, suppose you have four sample points within your plot and two answers (e.g. land cover class) to choose from. If one sample point is assigned to one answer and the other three points to the second answer, the data when downloaded will say '25' for the first answer and '75' for the second answer.

If you would like your plot data analyzed differently, the **Download Sample Data** option is a better fit. Downloaded .csv data from **Download Sample Data** will have the following columns:

1. **PLOT_ID**: the CEO-assigned unique sample plot number.
2. **SAMPLE_ID**: the CEO-assigned unique sample point number.
3. **LON** and **LAT** are the geographic coordinates of the center of your sample points.
4. **FLAGGED**: This will be FALSE for plots where data was collected and for plots where data has not been collected yet. It will be TRUE when a user has flagged the quality of the background map as not good enough to analyze the samples (e.g. due to clouds or poor image resolution).
5. **ANALYSES** will show 1 for plots that are analyzed, and 0 stands for not analyzed. If more than one person assigned a label to a point, this column will indicate the number of analysts that have assessed the point.
6. **USER_ID** is the user id (email address) of the person that classified the plot.
7. **COLLECTION_TIME:** The date and time when the user classified the plot. Timezone is UTC.
8. **ANALYSIS_DURATION:** Time in seconds that the user took to analyze the plot.
9. **IMAGERY_TITLE:** Name of the Imagery layer that the user had selected while analyzing the plot.

If multiple imagery sources were used, only the name of the last imagery layer used will be recorded. There is no way to know all the layers used by the user, e.g. if you want to check that two years of imagery were used.

10. If you used a .csv or .shp file for sample plot design, any additional data columns you uploaded will be preserved in the .csv download. They will be preceded by **PL_(column name)**.
11. If you used a .csv or .shp file for sample point design, any additional data columns you uploaded will be preserved in the .csv download. They will be preceded by **SMPL_(column name)**.
12. All the following columns will have information about each of the survey questions. They will be labeled **QUESTION TEXT**, where question text is the literal text of the question.

Note that imagery dates are not available as many of the imagery sources are composite. This means that each map tile is stitched together from imagery acquired on multiple dates. There is not a single date for an imagery tile.

If you are using SecureWatch imagery, you will have four additional columns:

13. **IMAGERYDATESECUREWATCH** will have a value for any samples which were classified while a specific date was selected from the imagery date dropdown.
14. IMAGERYSTARTDATESECUREWATCH, IMAGERYENDDATESECUREWATCH, and FEATUREPROFILESECUREWATCH will have values for any samples which were classified while a date range and feature profile were selected.
15. Note that imagery properties are associated with samples (not plots) because users are free to change these properties while classifying samples. Thus, any given plot may have some of its samples classified with one map image and other samples classified with a different map image.

F. Project statistics

If you click on **[Project Statistics]** you will see an overview of the number of members, contributors, total plots, flagged plots, analyzed plots, unanalyzed plots, and the date the project was created, published and closed.

These statistics overlap with the information shown on the **Project Dashboard**. Admins do not count towards the project's member count.

Part 8: Creating a Project for Validation

Creating a project in CEO to validate a spatial model—for example, a land cover map or a forest change map—presents some unique sampling and project design issues. In this section, we provide a very brief overview these issues both for single and for ensemble change maps.

Many resources provide a more in-depth review of this subject. Consider looking at:

1. CEO's Theoretical Manual, available here: https://collect.earth/downloads/CEO_Theoretical_Manual.pdf
2. the FAO's *Map Accuracy Assessment and Area Estimation: A Practical Guide* available online here: http://www.fao.org/3/a-i5601e.pdf
3. Documentation in AREA^2, available online here: https://area2.readthedocs.io/en/latest/background.html
4. A great overview on confusion matrixes here: http://spatial-analyst.net/ILWIS/htm/ilwismen/confusion_matrix.htm

A. Accuracy assessment

Accuracy assessments allow producers of spatial models to understand how accurate they are. This information is important for many reasons, including improving accuracy and determining if the product is suitable for management use.

For single models, such as a single land cover classification map, a confusion matrix is used to calculate the accuracy of spatial models. The purpose of this error analysis is to quantify two key metrics for each land cover class as well as the overall classification:

1. Given that a pixel is of a (land cover class, etc.), what is the chance that it was correctly classified as that (land cover class, etc.)?
2. Given that a pixel has been classified as a (land cover class, etc.), what is the chance that it belongs to that (land cover class, etc.)?

There are two pieces of information needed to answer these questions: the 'true' land cover class assignment, which is done by a human, and the 'as classified' land cover class assignment, which is output by the model. An example of the end goal (the confusion matrix) is shown below for a two class system.

	Actual: Forest	Actual: Not Forest
Predicted: Forest	29 (true positives)	22 (false positives)
Predicted Not Forest	12 (false negatives)	50 (true negatives)

For ensemble models, the approach is dependent on the type of ensemble model. For ensemble models where you have combined the results of the different models into a single model (e.g. through averaging), you can use an approach like that used for single models.

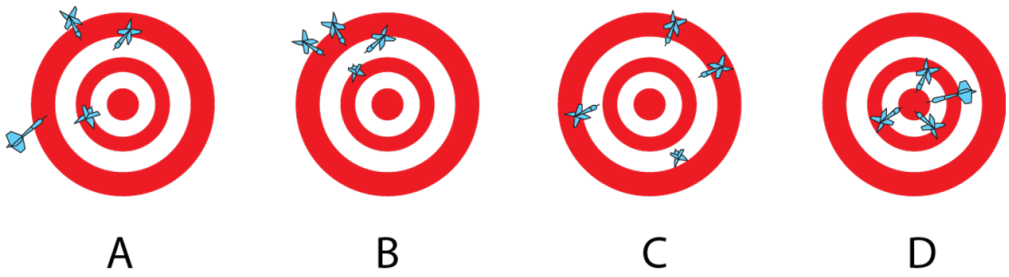
Where you have multiple different maps, you may need a more complex approach. For example, if you have two model outputs you may choose to stratify based on the values of both outputs. We will not discuss more complex approaches in depth here.

B. Validation goals & required sample size

Precision & accuracy goals for a project help us decide on the sample size and sample design.

Precision is how close your measurements or estimates are to one another. This is sometimes called dispersion. Precision describes random errors & variability.

Accuracy is how close your measurements or estimates are to the true answer. Accuracy describes systematic errors—including observational error. In statistics this is sometimes called bias instead of accuracy. Accuracy is sometimes hard to evaluate as the "true" value might not be known—but with Land Use & Land Cover analysis usually it can be.



Precise?	No	Yes	No	Yes
Accurate?	No	No	Yes	Yes

In this example, A is neither precise nor accurate. B is precise, but not accurate. C is accurate, but not precise, and D is both precise and accurate. This is of course where we would like to be ideally.

Accuracy and precision influence the necessary sample size. Detecting small changes in outcome (e.g. 4% decrease in forest cover) with high certainty requires large sample sizes.

When you are determining your sample size, you will want to consider the primary goals of the project, along with the resources available for data collection.

There are many different formulas for determining sample sizes. Which one you want to use is based in part on your sampling method (simple random sample, stratified random sample, etc.).

Two examples follow:

From SEPAL.io:

Formula to calculate the overall sample size

The equation below calculates an adequate overall sample size for stratified random sampling that can then be distributed among the different strata.

- N is number of units in the area of interest (number of overall pixels if the spatial unit is a pixel, number of polygons if the spatial unit is a polygon)
- S(O) is the standard error of the estimated overall accuracy that we would like to achieve
- Wi is the mapped proportion of area of class i
- Si is the standard deviation of stratum i.

$$n = \frac{(\sum W_i S_i)^2}{[S(\widehat{O})]^2 + (1/N)\sum W_i S_i^2} \approx \left(\frac{\sum W_i S_i}{S(\widehat{O})}\right)^2$$

From the CEO Technical Manual:

Expected accuracy of the product (P_0)
Precision of detecting differences from this accuracy (minimum detectable difference, δ)
Tolerance of Type I error (alpha, α)
Tolerance of Type II error (beta, ß)

$$n' = \left[\frac{Z_\alpha\sqrt{P_0(1-P_0)} + Z_\beta\sqrt{P_1(1-P_1)}}{\delta}\right]^2$$

Equation 1. The sample size determination equation.

The following equation can then be applied:

$$n = \frac{n'}{4}\left(1 + \frac{2}{n'\delta}\right)^2$$

Equation 2.The continuity correction.

Area Estimation & Accuracy Assessment (AREA^2) also has helper scripts for determining sample size, including assigning area-based weights. Instructions can be found here: https://coded.readthedocs.io/en/latest/sample.html.

If you have access, Foody (2009) also has a good approach specifically tailored to classification: see Foody, G. M. Sample size determination for image classification accuracy assessment and comparison. Proc. 8th Int. Symp. Spat. Accuracy Assess. Nat. Resour. Environ. Sci. 30, 154–162 (2008).

C. Sampling approach

For validating a single model or an ensemble model that has been averaged/combined, stratified sampling is the preferred method. This is to ensure you can accurately and precisely determine how well your classification performs for each of your predicted groups (land cover classes, forest degradation, etc.). If you use random or gridded sampling, you might not sample enough points in rare classes to be able to tell if they are well estimated.

Strata used for stratification must be exhaustive--they include the entire study area--and they must be exclusive--a sample unit can't belong to more than one strata.

CEO cannot currently create stratified sampling designs. However, you have multiple good options to create a stratified sample outside of CEO.

1. You can use SEPAL.IO to create a stratified sample based on how accurate you think your model is for each stratum and a raster of your model output.
2. You will need to sign up to use it by visiting sepal.io and clicking on **[Sign Up]**.
3. From the apps, find the **Stratified Area Estimator-Design** tool.
4. From there, follow sepal's directions.
5. You can also run this on your own machine; see the instructions here: https://github.com/openforis/accuracy-assessment

You can also use QGIS or ArcGIS for creating a stratified sample. Many resources are available online, including:

- This blog post for QGIS: https://pvanb.wordpress.com/2012/12/03/stratified-random-sampling-in-qgis/ Note you will need to convert your classification raster to a vector layer.
- This description of the Sampling Design Tool in ArcGIS: https://www.arcgis.com/home/item.html?id=28f08ca526ae44e8ac107a2a0d5f50e3

Once you have created a stratified sample, you can create a project in CEO and follow the directions in Part 4: H Uploading CSV & SHP files to upload your sampling points into CEO.

After collecting data on all of your sampling points, you can download the data (see Part 7: E Download your data) and calculate your confusion matrix.

D. Alternative: Using TimeSync

TimeSync is an application that allows researchers and managers to characterize and quantify disturbance and landscape change by facilitating plot-level interpretation of Landsat time series stacks of imagery (a plot is commonly one Landsat pixel).

TimeSync can be used to validate map data products, especially where estimating change is a primary objective. Also, TimeSync can facilitate production of independent estimates of change and disturbance rates from remotely sensed imagery.

CEO is currently integrating TimeSync functionality into its web platform. More information on TimeSync functionality in CEO will be available when this integration is complete.

For more information on TimeSync please see: http://timesync.forestry.oregonstate.edu/. A tutorial on how to use TimeSync is available at: http://timesync.forestry.oregonstate.edu/tutorial.html.