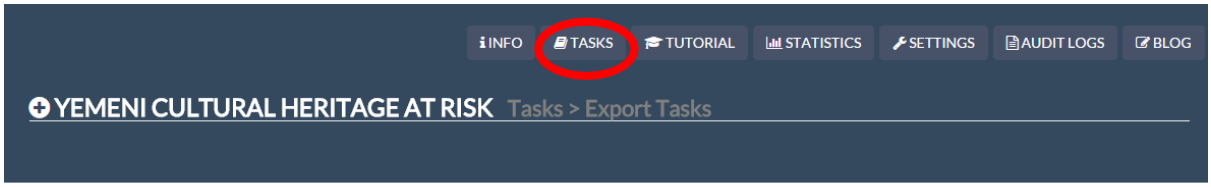
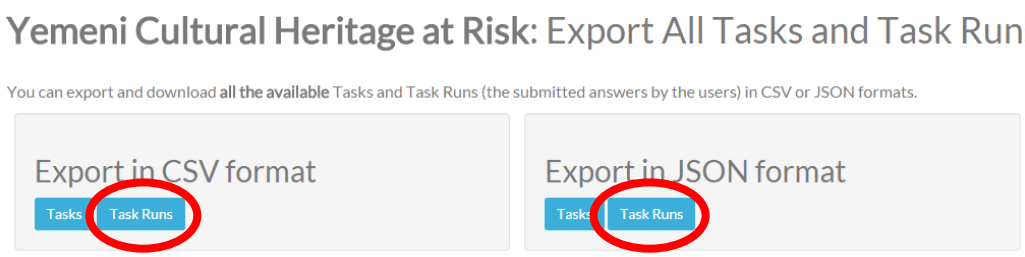
How can I get hold of GeoTag-X datasets, and what do they tell me ?

Results from Geotag-X can be downloaded in a variety of formats. For the rest of this document when we refer to "Tasks" we mean the photos that are being shown for analysis:

1) Raw datasets

These can be downloaded from within the projects via the "Tasks" link and "Export Tasks"





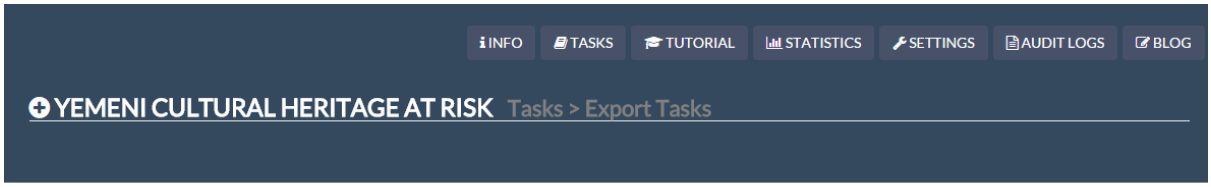
Raw data are available in both CSV and JSON format. To download the full list of tasks being shown for analysis, you want to click on the link to download "Tasks". If you are interested in results you want to click on "Task Runs". Both CSV and JSON contain the same data, though in slightly different ways. The following table explains what each field is for the Yemen Cultural Heritage at Risk project. Any field that begins with *task_runinfo* contains the answers to the questions for that project and will be specific to that project:

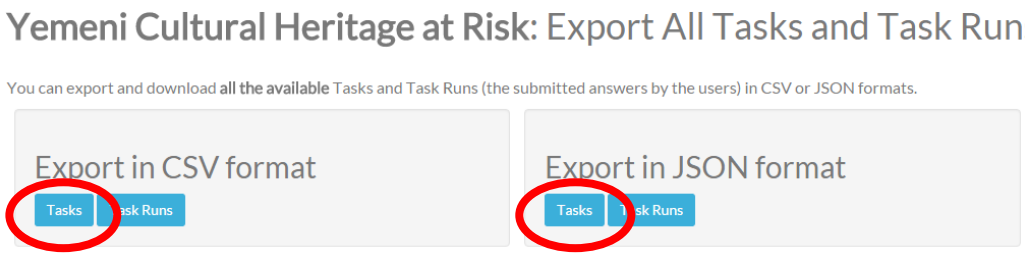
CSV	JSON	Description
Task_run_calibration	calibration	Whether or not the analysis has been calibrated against a correct answer. This will always be *null* because GeoTag-X does not use this function in PyBossa.
task_run_created	created	This is the exact time at which this entry in the database was created
task_run_finish_time	finish_time	This is the exact time at which the analysis was finished (currently this is the same as the *created* time)
task_run_id	id	This is the unique id for this particular analysis.
task_run_info		This is all of the answers given by the analyst. This corresponds to the *Info* group of fields in the JSON
task_run_project_id	project_id	This is the unique numerical identifier of the project
task_run_task_id	task_id	This is the unique numerical identifier of the task that was analysed
task_run_timeout	timeout	Used internally by PyBossa
task_run_user_id	user_id	The numerical ID of the user who completed the analysis. This cannot be used to identify the user by someone downloading the data. This will be *null* if the volunteer was anonymous.
task_run_user_ip	user_ip	This is the IP address of the analyst. This is only recorded when the analyst was anonymous, otherwise it will be recorded as *null.*
task_runinfo_damage	damage	This is the answer the volunteer gave to the question "Do you see any damage to the structure that is possibly not caused by general deterioration over time?"

task_runinfo_geolocation	geolocation	This is where the volunteer thought the photo has been taken. The numbers are the coordinates in latitude and longitude for the vertices (corners) of the polygon drawn by the volunteer in the geotagging question.
task_runinfo_heritage	heritage	This is the answer the volunteer gave to the question "Do you think there is built heritage in this photo?" This is the first question of the application and is used to determine whether or not the photo is relevant to the rest of the analysis.
task_runinfo_img	img	This is the url to the photo (task) being analysed
task_runinfo_name	name	This is what the volunteer though the site in the photo was called. This comes from the question "What is the name of the site shown in this photo?"
task_runinfo_wikipedia	wikipedia	This is the Wikipedia page that the volunteer found for the site in the photo. This comes from the question "What is its wikipedia page?"

2) List of tasks

On the same download page you can download a list of tasks being presented for analysis:





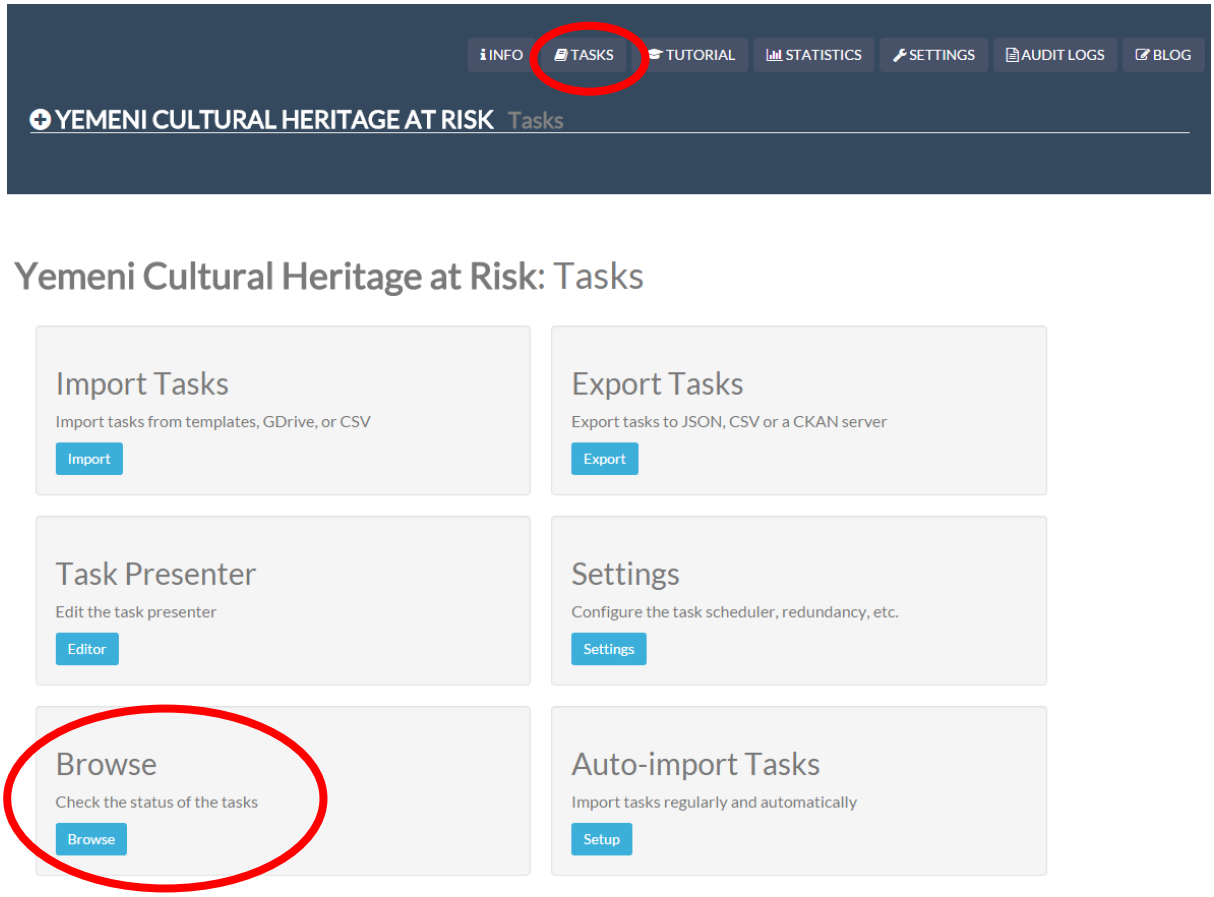
The following table explains the fields in the tasks download:

CSV	**JSON**	**Description**
task_calibration	calibration	Used internally by PyBossa
task_created	created	The time and date that the task was added to the project
task_id	id	The unique ID for this task
task_info	info	All the data collected about the task. The same data as that under the *taskinfo* fields
task_n_answers	n_answers	Used internally by PyBossa
task_priority_0	priority_0	Used internally by PyBossa
task_project_id	project_id	The numerical id of the project from which the tasks were downloaded
task_quorem	quorem	Used internally by PyBossa
task_state	state	Whether or not the task has been completed. Completed means that the required number of analysts have analysed the task. Ongoing means that the task is still being presented for analysis
taskinfo_id	info id	Used internally by PyBossa
taskinfo_image_url	info image_url	The direct URL to the photo on the external website where it is hosted. This is used to display the photo in GeoTag-X

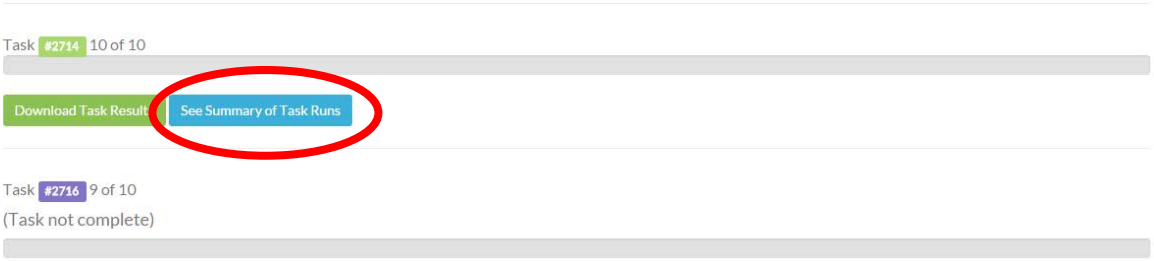
taskinfo_source_uri	info source_uri	the URL of the source of the photo

3) Aggregated data visualisation per task

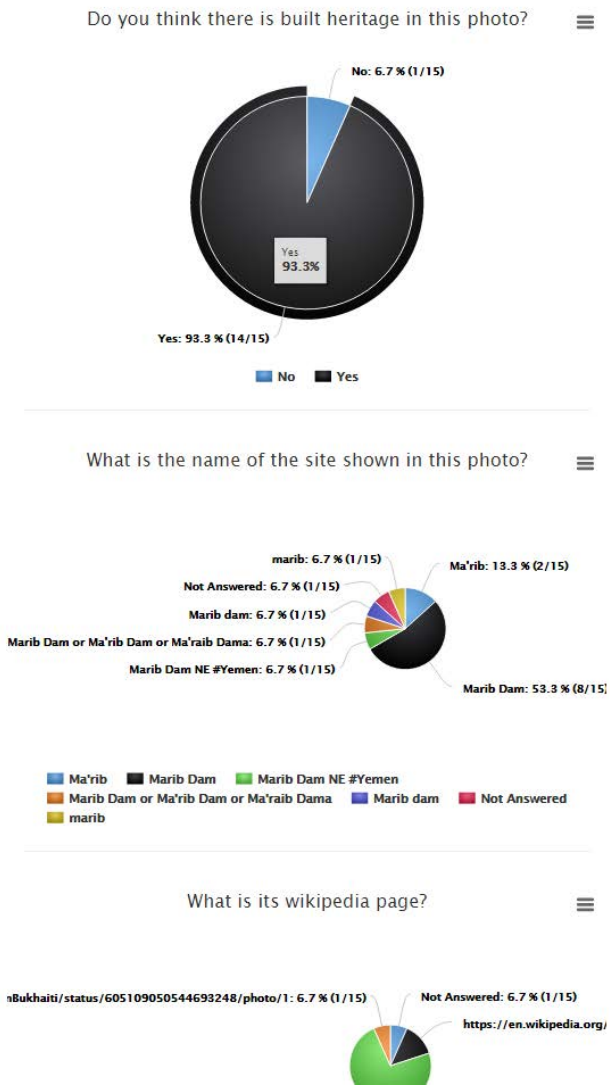
We have aggregated the results and provide a pie chart visualisation for each task. You can find this under the “Task” link for each project and then “Browse”:



If a task is complete, meaning it has had the required number of analysts analyse it, you will have the option to see the aggregated results. To see the results for any particular task that has been completed you just need to click on “See Summary of Task Runs”. If the task has not been completed this option will not be available:



You will see the photo and a set of pie charts showing the percentage of people who gave which answers:

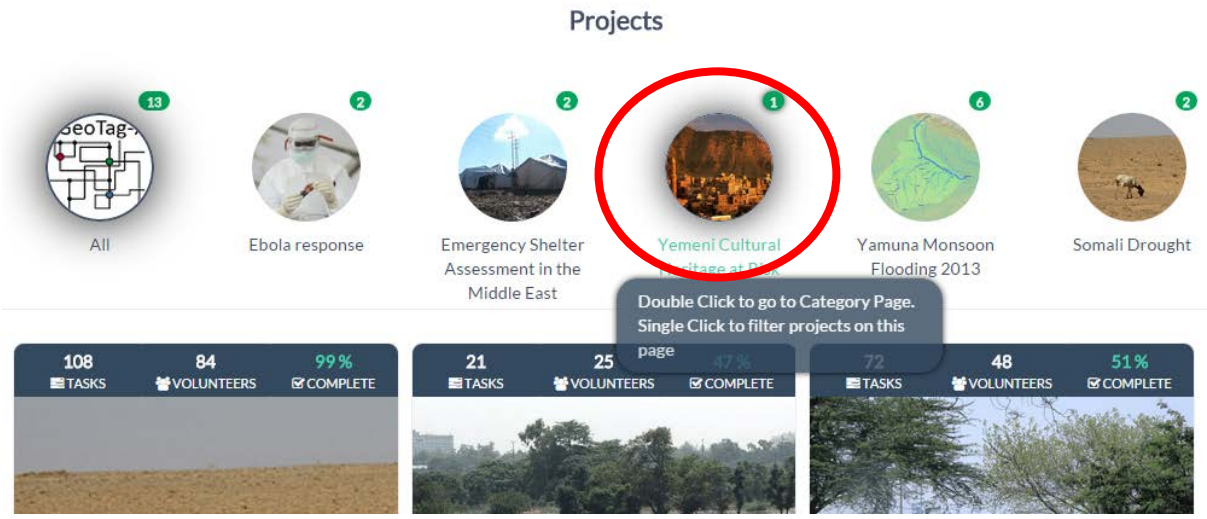


For instance in this example from Yemeni cultural heritage at risk, the first pie chart shows that 93.3%, or 14 out of 15 analysts, said that there was built heritage in this photo. The second pie chart shows the different answers the analysts gave as the name of the site.

You can also download the raw data for only that particular task under “Download Task Results”. This is provided in JSON format and is an extract of the complete raw data as described in the previous section so the fields will be the same as for the complete raw data download (table 1).

4) Aggregated data GeoJSON

We also provide the aggregated data for each category in GeoJSON format available for download from the categories page (double click on the category icon on the homepage):



The database of tasks is associated with a category, and each task will appear in all projects under a given category. The GeoJSON contains the results from all projects for each task. Here are the fields explained:

```
{
  "features": [
    {
      "geometry": {
        "coordinates": [
```

Geometry and coordinates: the list of coordinates for each polygon as defined by the volunteers when they geotagged the location of the photo. This section can contain data for multiple polygons, depending on how many volunteers geotagged the photo.

```
},
"properties": {
  "GEOTAGX_IMAGE_URL": "https://farm9.s
  "reach_geo::GEOTAGX_TOTAL": 19,
  "wintershelter::GEOTAGX_TOTAL": 30,
```

properties contains the image URL (*GEOTAGX_IMAGE_URL),* and the number of volunteers who analysed the task in each project. This example comes from the Emergency Shelter category. The GeoJSON tells us that 19 people analysed the task in the geotagging project (*reach_geo::GEOTAGX_TOTAL)* and 30 people analysed the task in the "are shelters prepared for winter" project (*wintershelter::GEOTAGX_TOTAL).*

```
"wintershelter::shelter": {
  "answer_summary": {
    "Yes": 30
  },
  "question_text": "Do you see shelter in this photo?"
},
"wintershelter::shelterchimney": {
  "answer_summary": {
    "No": 6,
    "NotClear": 4,
    "Unknown": 19,
    "Yes": 1
  },
  "question_text": "Is there space to put a chimney safely inside the shelter?"
},
"wintershelter::shelterflysheet": {
  "answer_summary": {
    "NotClear": 1,
    "Yes": 29
  },
  "question_text": "Does the shelter have a second cover to protect it from the rain?"
},
```


Properties also contains the data for each question in the analysis. In this example from the "Are shelters prepared for winter" we can see the results for three questions. For each question, the first line shows the project that it came from and the short identifier for that question (eg *wintershelter::shelter, wintershelter::shelterchimney, wintershelter::shelterflysheet*). The second section, *answer_summary* contains all of the responses given by the analysts, and the number of analysts who gave each response. The final line, *question_text* gives the exact question that was asked in the analysis. So for the first question, we can see that the question asked was "Do you see shelter in this photo", and that all analysts answered yes (30 out of a total of 30 analysts). In the second question we can see that the question was "Is there a space to put a chimney safely inside the shelter?" and that 6 analysts said "no", 4 said "not clear", 19 said "unknown", and 1 said "yes".

How do I get the GeoJSON into a GIS?

The GeoJSON can be opened in QGIS, a free, open source GIS. You can either work with the data in QGIS, or export it as a shapefile that can then be opened in ArcGIS or another GIS software.

5) Map

The GeoJSON can be visualised on the provided map (where the tasks for the given category have been geotagged).

YEMENI CULTURAL HERITAGE AT RISK

The on-going civil war in Yemen is putting much of its cultural heritage at risk. With this project we are looking to collect as many photos of cultural heritage sites as possible, tagging them with their name and any relevant wikipedia entry, and then geotagging them as best as possible

Total Number of Projects: 1

FIND PHOTOS FOR THIS PROJECT | VISUALIZE SUMMARY ON MAP | EXPORT SUMMARY AS GEOJSON