Vera Rubin Observatory/Zooniverse

Citizen Science Project overlap policy

Chris Lintott, April 2024

1. Motivation:

Citizen science projects such as those hosted on Zooniverse depend on the goodwill and participation of volunteers to produce results. We know from surveys (e.g. Dowthwaite et al 2019, SSRN, <https://dx.doi.org/10.2139/ssrn.3480452>) that a primary motivation is a desire to contribute to research; it is therefore also in the interest of the volunteers, as well as the project, that sufficient classifications are obtained by a project. Communities of interest (see Luczak-Rösch et al 2014 <https://eprints.soton.ac.uk/363523/> for a discussion) also grow around long-lived projects such as Galaxy Zoo or Active Asteroids, and these play a vital role in providing volunteers with opportunities for skill development and learning.

It is therefore desirable that, where possible, we avoid a situation in which the platform hosts multiple, competing projects which share science cases. This is particularly important as it would be difficult for volunteers to judge which project teams are best positioned to use their classifications (a process which depends on factors such as available effort for research).

While noting that there is no desire amongst the Rubin citizen science team, nor among the science collaborations to enforce a ‘no compete’ policy, we believe most participating researchers would welcome the chance to collaborate, especially if overlapping project interests were identified early in the process of project building. This policy aims to aid in identifying and alerting such researchers to potential overlaps.

2. Technological solution

The original intention was that the platform could be proactive in warning researchers of potential overlapping projects. As users will use the tools built in the RSP (Rubin Science Platform) to connect to the Zooniverse API and transfer data, there exists the potential to check for projects which are using datasets which substantially overlap with each other.

Such a system should intervene as early as possible in the process of project building; once teams have invested substantial time in building projects they are less likely to be willing to join forces. An ideal system would alert users to an overlap before they have built a project.

The citizen science EPO team had previously implemented a system for tracking data use by researchers transferring images to the EPO cluster (it was originally anticipated that images used by citizen science projects would need to be served from this cluster). Intending to use a simple metric (e.g. number of objects in common between datasets), we assessed the possibility of using this system to check on the point of transferring data to Zooniverse whether a threshold percentage of subjects had been used before. If so, the user could be provided with the identity of the owner of the project, and the project status (live/in preparation/complete).

However, it was realised that the decision made not to provide consistent object IDs across data releases (see document LSE-163 sec. 2.4) would mean that this could not be a simple search across IDs. While a proximity search with coordinates is possible, this would be resource heavy (particularly once several projects involving large sets of data were launched) and would impact performance of the data transfer service significantly.

In coordination with the EPO team, we decided not to pursue a technological fix, but rather to develop a process of proactive communication about project development.

1. Policy and communication solutions.

Rubin data rights holders are able to transfer a small (~few hundred) number of images to the Zooniverse platform without any explicit permission. This allows for the development of projects on the platform using real Rubin data. Following project development, the team apply for permission to run a beta test (from the Zooniverse team) and simultaneously for the release of the larger dataset required for the project (from Rubin).

2.1 At this point the LSST:UK Zooniverse team will:

* Inform the team, ahead of beta, of any existing Zooniverse/Rubin projects in similar science areas.
* Check the requested data release against data already released for other projects, in case this identifies substantial overlap (for example, it may make sense for projects both studying the same area of sky but with different science goals in mind to collaborate)
* Notify a contact identified by the science collaboration(s) of the existence of a new project.

2.2 Projects which make use of annual data releases are likely to be prepared early in the release cycle. In the lead up to each data release we will:

* Contact existing Rubin citizen science project leads to remind them to look for opportunities to collaborate.
* Contact science collaborations to encourage them to coordinate citizen science efforts where possible.
* Provide to those with data rights an updated directory of ongoing Rubin citizen science efforts, including those which are under construction. This will need to be populated by project owners, and cannot be made compulsory, but should help with project discovery.

1. Broker projects.

Projects using transient alerts accessed via the community brokers have special status, as their owners can pass data to Zooniverse without going via the RSP. Broker projects will still have to apply for review and so the steps outlined in 2.1 above will still apply.

In order to help understand overlap between projects, we will ask project owners to supply the queries used to select data in each case as well as the standard information.