Table A1. Citizen Science Data Guide to Coding (adapted from Theobald et al., 2015)

To be included, all programs met the following two definitions:

\* Definition of Citizen Science (CS) Programs: Volunteer(s), who may or may not be experts, collect and/or process data as part of scientific enquiry.

\* Definition of Biodiversity Data: Information on the numbers (or presence/absence) and identifies a taxonomic group (e.g. species, genus, family). Must include information on space and time of the samples collected.

\* Selection of Biodiversity CS programs to be included:

Programs were prioritised if they are already (or likely to be) contributing biodiversity data for monitoring purposes, where the data are quality controlled and shared—e.g. verification of the data collected, specific verification techniques listed, etc.

CS programs were also considered if they collected data at scale (more than 100 observations were collected per year).

Some programs were included that did not meet some of these requirements if they were deemed unique representations covering an underrepresented EBV, taxon, or geography. We included a number of projects that are no longer supported provided that they collect data at scale.

Further information about each column

|  |
| --- |
| Program Info |
| **Program Name:** the name of the CS program |
| **Goal Statement:** the program’s self-identified goal. A quoted mission statement or other statement describing the goal of the program. Sometimes paraphrased or interpreted if no clear statement was given. |
| **Portal:** (Yes/No) yes if the website hosts more than one separate CS project |
| **# projects** supported by the online CS program |
| **Program Headquarters Type:** headquarter type of program (Options: Governmental, NGO, Academic, Other, Multiple, Unknown) |
| **Headquarters Continent:** continent where program is headquartered (North America, South America, Europe, Africa, Asia, Oceania, Antarctica) |
| **Start Year:** year the program was established |
| **End Year:** year the program was terminated (if no longer running) |
|   |
| Data Info |
| **Type of sampling:** Extensive (maximising geographic coverage at the expense of sampling effort per site, as expressed as the number of ecosystem variables or functional groups monitored and/or sampling frequency); Intensive (maximising sampling efforts at a site at the expense of sampling more sites); both (taken from Proenca et al 2016) |
| **Data Collection Continent:** location of data collection, continental scale (North America, South America, Europe, Africa, Asia, Oceania, Antarctica, global (more than 1 continent)) |
| **Data Collection Ecosystem:** Terrestrial, Marine, Freshwater, Estuarine, Multiple |
| **Taxa:** the focal organism(s) of the program. As detailed as possible. Use Other if no specific organisms identified (e.g. water quality, soil moisture) |
| **Invasive:** if the focal organisms are invasive (y/n/some) |
| **Essential Biodiversity Variables (EBV) -** types of EBV data collected |
| **Data Verifiable:** Are data (e.g. taxonomic identifications) verifiable? In other words, is there some mechanism in place to check the data's accuracy? Options: All (if all data can be verified), Some (if only some data can be verified, for example if photograph available) |
| **Data Verification Mechanism:** If Yes in previous column, what is the mechanism? Options: photo, specimen collection, expert present, combo, other, NA (If None or Unknown for Data\_Verifiable) |
| **Data Readily Available:** Are the data available publicly online? Options: All (if all data collected are available, i.e. in a spreadsheet or other detailed info), Some (if some of the data collected are available, such as on a map or summarised in a figure or report |
| **Pubish to GBIF.org** directly or indirectly via another portal (e.g. ALA) or repository  |
| **Sample Standardised:** Is data sampling standardised? Options are ALL (data are collected whether or not organism is found; effort-time or area- is quantified), SOME (some data sampling is standardised; this is for programs that have two or more different types of sampling |
| **Participant - Community focus:** Is project designed to engage local community members to monitor their local environment: Y- main focus is to engage local community members, with project goal tied to local issues, S - some of focus is targeted to local community |
| **Program Website:** the web address for the program |