



This project is co-financed by the European Union

Grant Agreement No.: 824603  
Call: H2020-SwafS-2018-1  
Type of action: RIA  
Starting date: 1/02/2019



# **Lifecycle-aware citizen science templates**

**Coordinator: Esteban Gonzalez, Oscar Corcho**  
**Quality reviewer: Neal Reeves**

<b>Deliverable nature</b>	Technical report
<b>Dissemination level</b>	Public
<b>Work package and Task</b>	WP4 T1.2
<b>Contractual delivery date</b>	30/01/2020
<b>Actual delivery date</b>	

**Authors**

<b>Author name</b>	<b>Organization</b>	<b>E-Mail</b>
Esteban Gonzalez	UPM	egonzalez@fi.upm.es
Óscar Corcho	UPM	ocorcho@fi.upm.es

<p><b>Abstract</b></p>	<p>This document describes the different templates that are going to be developed in ACTION for helping pilots to export/use external platforms. Also, a new tool to create Data Management Plan documents based on a questionnaire will be described. Finally, a mini guide has been included to help users to create a CS project using the external platforms Epicollect and Zooniverse.</p>
<p><b>Keywords</b></p>	<p>Citizen science platforms, data management</p>

**Disclaimer**

*The information, documentation and figures available in this deliverable, is written by the ACTION project consortium under EC grant agreement 824603 and does not necessarily reflect the views of the European Commission. The European Commission is not liable for any use that may be made of the information contained herein.*



*This deliverable is licensed under a Creative Commons Attribution 4.0 International License*

**How to quote this document**

*Gonzalez, E. (2020). Lifecycle-aware citizen science templates*

## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY</b>	<b>4</b>
<b>1 INTRODUCTION</b>	<b>5</b>
<b>2 EXTERNAL PLATFORMS</b>	<b>6</b>
2.1 Zooniverse	6
2.2 SciStarter	8
<b>3 TEMPLATES</b>	<b>9</b>
3.1 Introduction	9
3.1 Initialization template	9
3.2 Data Management Plan templates	10
3.3 Data Management Workflow templates	12
3.4 Presentation template	15
<b>4 DMP creator tool</b>	<b>17</b>
<b>5 CONCLUSIONS AND NEXT STEPS</b>	<b>20</b>
<b>ANNEX I: Mini tutorial to create a CS project using Epicollect and Zooniverse</b>	<b>22</b>
<b>ANNEX II: Original set of questions in H2020 DMP template</b>	<b>32</b>

## EXECUTIVE SUMMARY

In the context of our project, templates are mechanisms to build citizen science projects following the *good practices* that will be analyzed in ACTION, in particular in WP5. These templates can be either tools (software) or guidelines to facilitate the use and the integration of tools in Citizen Science projects, covering the steps of their lifecycle.

The initial objective of this deliverable is to create templates to *export* projects to the most popular open citizen science platforms such as PyBossa, Zooniverse and SciStarter. We will see in the following sections, the use of these platforms lay out some problems (the closure of the platform Crowdcrafting, PyBossa service ), and SciStarter (impossibility of creating projects based on tasks).

After an analysis made in deliverable D4.1 (Conceptual Architecture) about citizen science projects, we discover that many of the projects suffered a lack of data management and policies.

Due to these facts, we decided to expand and group the templates in four categories:

- The objective of the first template is to create projects in the Zooniverse platform. Basically, this template will set up a project in Zooniverse, with some of the main concepts exposed in deliverable D5.1(Task design)
- A second set of templates designed to create DMP (Data Management Plans) specifically designed for citizen science projects following the FAIR<sup>1</sup> principles.
- A third set of templates to build simplified scientific workflow, focused on citizen science projects, to allow users to manage the data generated in their projects. This workflow will follow the plan exposed in the DMP associated and thus, following the criteria of FAIR principles.
- The last one has the objective of disseminate the objectives of results of a CS project. This template is an HTML webpage with a similar structure for all pilots they decide to use it. Data and results will be added automatically by the other tools that will be generated in the project (see D4.1).

This deliverable exposes, in this first version, the progress made in each template, and the next plans.

---

<sup>1</sup> <https://www.go-fair.org/fair-principles/>

## 1 INTRODUCTION

The initial objective of this deliverable as was described in the project's proposal is to create a set of templates to create projects in external platforms such as Zooniverse, PyBossa and SciStarter. We will see in the next sections how we want to extend the functionality of these templates to other domains such as data management or dissemination.

In section 2, the functionality of some external citizen science platforms (PyBossa, Zooniverse and SciStarter) will be investigated in order to analyze how to integrate the templates with them.

In section 3, the different templates that are going to be developed in the project will be presented.

In section 4, a new tool to create Data Management Plan (DMP) will be described together with a mini guideline of use.

In section 5, the main milestones of the deployment of templates will be listed.

## 2 EXTERNAL PLATFORMS

The initial purpose of the templates is to *export* citizen science projects based on pollution domain to the main citizen science platforms. In this way, we can raise the participation and outreach of the project.

In this section, we are going to explain the different options than the external platforms offer, and how we can interact with them.

During the execution of the project, one of the citizen science platforms, Crowdcrafting<sup>2</sup> (based on PyBossa software), has closed its service as can be seen in the following tweet and visiting its webpage.



Fig 1. Crowdcrafting shutdown service message

### 2.1 Zooniverse

Zooniverse<sup>3</sup> is one of the most popular citizen science platforms. In this platform, citizens can create their own citizen science projects and feed them with tasks (built over images mainly). The platform is highly configurable and customizable<sup>4</sup>. In this analysis, we will focus on those parts that we will be filled in through the templates.

There are three main categories of parameters:

- **Definition of the project:** In this area, the project leader defines the objective and a brief description of the project. Also, more parameters such as introduction text, research question, tutorials and logos can be defined.
- **Workflow:** In this area, the coordinator or an authorized user can design the tasks that will be presented to the users. These tasks can be based on four types: questions, drawings, surveys and transcriptions. These tasks can be combined forming a workflow.
- **Subject Sets:** Once the workflow has been defined, this will be fed with the different tasks (mainly with images).

<sup>2</sup> <https://scifabric.com/crowdcrafting/>

<sup>3</sup> <http://www.zooniverse.org>

<sup>4</sup> <https://help.zooniverse.org/getting-started/>

### Definition of the project

In this part, the project leader defines information related to the description of the project. This information is used to create the project on the platform. This information can be used in the future to feed some templates/tools, for example, to describe the metadata needed to integrate its results in external repositories such as Zenodo or Github or to create websites for the projects.

We can highlight the following fields:

- Name
- Description
- Introduction
- Tags
- Social links

### Workflow

A workflow is a sequence of tasks that the volunteers execute to report observations. In this section, the project coordinator has to design the workflows configuring some fields. This includes:

- Title of the workflow
- Subject sets
- Tutorials associated to the workflows
- An option for the volunteers to see classification summaries
- An option for the volunteers to pan and zoom on the image
- An option to show a set of images instead only one image
- The limit of classification per subject

Besides these configuration fields, other fields can be shown up depending on the type of the task selected such as questions, drawings, surveys and texts.

In the case of **questions**, the designer can define a question and a set of possible answers that volunteers have to answer. This question is related to an image that was previously uploaded in the subject sets section. The designer can add the option to multiple answers..

In the case of **drawing**, volunteers can draw into the image presented different shapes such as circles, polygons, ellipses. The object is to create a region of interest for other tasks. This information is included as a result

In the case of **surveys**, volunteers identify objects in the image(s) by filtering by their visible characteristics, and then they answer some questions previously defined.

In the case of **texts**, volunteers write in a free field of text the transcription of a text that they are seeing in an image.



We will see in section 3 how this approach can be exported to Zooniverse.

### Subject Sets

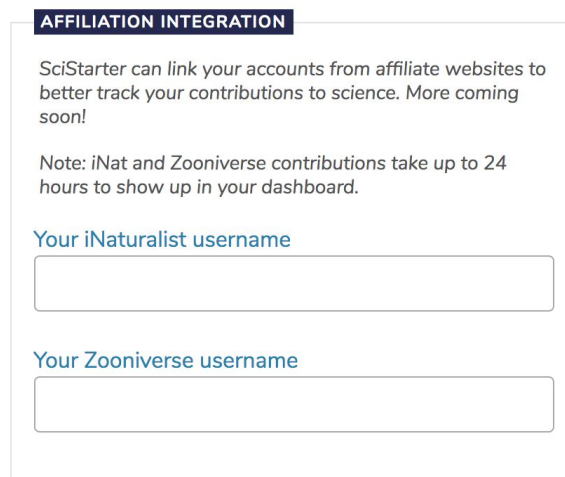
These are the resources (images) that can be classified by the volunteers. This information can be manually uploaded to the platform or automatically through the API or the panoptes application. Zooniverse has several clients developed in the most popular languages, to interact with its system. It allows external systems to build and configure a project, as well as feed the project with new tasks and images. It is the way that our templates will interact with Zooniverse.

## 2.2 SciStarter

Scistarter is a platform where users can register CS projects, and use different tools to track contributions of citizens as well as have access to a homogeneous login system based on OpenID.

As a first step, users have to fill in a webform that will be checked by Scistarter support team in order to register the project. This information can only be provided from a form that is available in the website<sup>5</sup>. Once the project is registered, the API (Participation API) to register contributions can be used.

Zooniverse and iNaturalist platforms can be registered in SciStarter to automatically send the user contributions to scistarter. Users only have to add their zooniverse and inaturalist usernames in the settings panel. This can be done easily from the settings tab in SciStarter (Fig. 1).



**AFFILIATION INTEGRATION**

SciStarter can link your accounts from affiliate websites to better track your contributions to science. More coming soon!

Note: iNat and Zooniverse contributions take up to 24 hours to show up in your dashboard.

Your iNaturalist username

Your Zooniverse username

Fig 2. Affiliation window between Zooniverse and SciStarter

<sup>5</sup> <https://scistarter.org/add-project>

## 3 TEMPLATES

### 3.1 Introduction

In the previous section, we have analyzed how Zooniverse and Scistarter platforms work and how users or other systems can interact with them. While Zooniverse is focused on designing tasks to analyze images, Scistarter is focused on publishing a project in a common platform (to engage newcomers), and to collect some metrics about users contributions.

Based on this, and the lack of data management in CS projects - identified in D4.1 Conceptual Architecture - we have split the templates in different groups according to the functionalities they adopt.

The next figure depicts the different groups formed to group the templates. Depending on the application environment, templates will be skeletons to generate other files or documents, or scripts to execute tasks. In the next subsections, each template will be described.

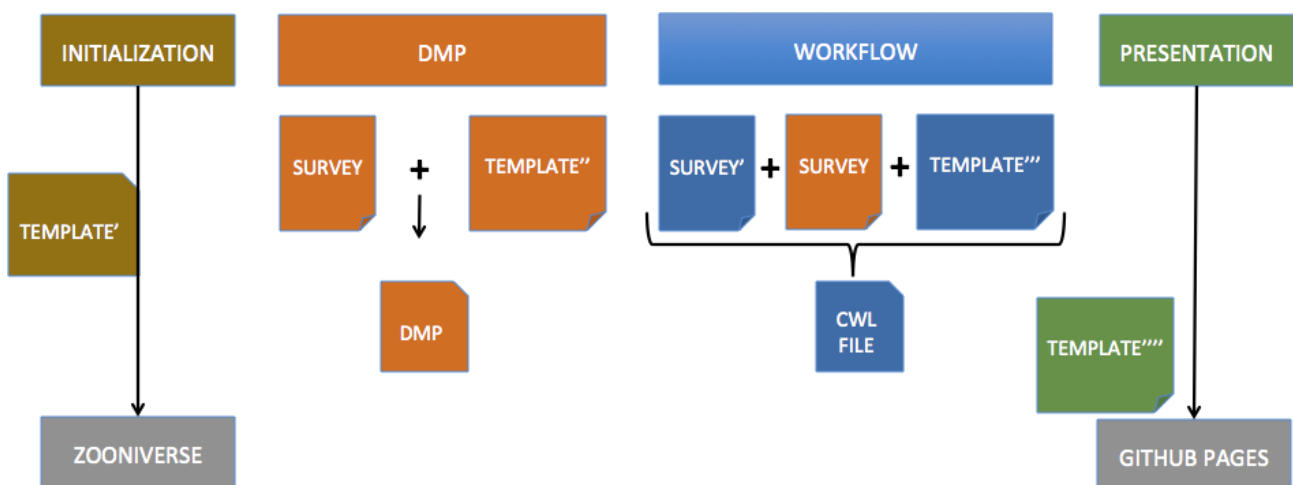


Fig 3. Categories of templates

### 3.1 Initialization template

This first template is used to initialize the projects to create projects in the platforms selected. As we have seen in section 2.2, Scistarter doesn't have a mechanism to automate this process, like an API, so for this case, there is no sense to develop a template for this action.

In the case of Zooniverse, we have to introduce some fields to configure a project with detailed information about the description of the projects. It includes the following information:

- **Name** of the project
- **Description** of the project

Lifecycle-aware citizen science templates

- **Introduction text**
- **Tags** to index your project in zooniverse platform.

This information will be defined in our template and exported to Zooniverse. In this case, we have prepared a small guideline, with an example, based on one of our pilots, to describe how to create a small CS project in Zooniverse.

### **3.2 Data Management Plan templates**

In modern projects, especially In H2020, a Data Management Plan has been delivered in order to describe the data that you are going to collect and how you are going to manage it. For example, in ACTION, a deliverable<sup>6</sup> has been sent to explain how the pilots are going to deal with their data.

But, what happens in CS projects? Are the citizens prepared to write these documents? After our experience in the open call process, designed to incorporate 6 new pilots, we can say no.

If you look at a DMP template, it is designed to be filled in by an expert on data management. But, in the case of pollution-related CS projects, they have a similar structure, differentiating in some parameters that can be configured.

For this reason, we have decided to create homogeneous pollution-related DMP (based on H2020 projects) for all ACTION-supported projects. What are the ingredients we need to build this template?

- First, we need a template with empty fields that can be filled with the configuration provided by users. We have chosen the schema Docbook<sup>7</sup> because it is easier to export to another popular format such as PDF or HTML. Plus, the elements of the document can be tagged with concepts that can be easily processed by external modules.
- Second, we need a webform to allow users to configure the tool with the information needed to fill in the document.
- Third, we have to transform the sections of the DMP in a questionnaire with direct questions and controlled answers.

As a result of applying these points, ACTION has developed a tool to facilitate the creation of these DMPs. This tool will be described in the next section.

The crucial part of this tool is the elaboration of a simplified questionnaire to generate the DMP. For that, the following process has been applied:

1. We start from the questions contemplated in the official H2020 DMP template documentation (see Annex II).

---

<sup>6</sup> D3.1 Data Management Plan

<sup>7</sup> <https://docbook.org/whatis>

Lifecycle-aware citizen science templates

2. We remove the questions related to the application of FAIR principles because it will be added by us automatically. Users always can manually edit the document to satisfy the requirements of their project.
3. We rephrase some questions in order to make them more understandable to non data expert users.
4. Finally, we show the questionnaire to our CS pilots in order to get feedback from them.

There is a list of 40 questions that every project has to fulfil. This list of questions can be seen in Annex II.

In order to simplify the questionnaire, we make the following assumptions:

1. The data will be managed from the tools or services that ACTION provides (see deliverable D4.1 - Conceptual Architecture).
2. Data will follow the Open Science and FAIR principles.
3. Data will be published in Zenodo generating the appropriate metadata and versions.

As a result, the following questionnaire has been designed and implemented in our tool.

Activated with option	Question	Response Type (format)	Options
	What is the purpose of the data collected/generated?	Free text	
	Could you provide us a description of the data you are going to generate?	Free text	
	Will you re-use any existing data? (have you used existing data?)	Choices	[Yes, No, I don't know]
Yes (5)	Please, specify where it can be found	url	
	At this moment, is your project collecting data?	Choices	[Yes, No]
Yes (7)	Could you specify the link of the data?	url	
No (7)	Do you want that ACTION publish and share your data on behalf you?	Choices	[Yes, No]
Yes (9)	Could you provide us some keywords that describe your data?	List of words	
No (9)	If not, do you think in future you'll change your mind?	Choices	[Yes, No]
Yes (11)	Could you establish a date to release the data, please?	Date	
No (11)	Explain the reason	Free text	



	Please, Specify the license of your data	Choices	[CC BY, CC BY-SA, CC0, CC BY-NC, CC BY-NC-SA, CC-BY-ND, CC BY-NC-ND, Other]
Other (14)	Please, write the terms and conditions to use your license	Free text	
	Does your data follow an specific vocabulary or standard to describe it?	Choices	[Yes, No, I don't know]
Yes (16)	Provide the vocabularies, standards or methodologies	Free text	
	Are you using any methodology or process to assure the quality of the data?	Choices	[Yes, No, I don't know]
Yes (18)	Please, describe it	Free text	
	Are you using personal information in your data (name, emails, telephone, etc ...) ?	Choices	[Yes, No]
Yes (20)	Named if you apply an anonymization process to the data	Free text	
	Are you using sensitive data for protected species (as geolocation)?	Choices	[Yes, No]
Yes (22)	Describe the process to avoid this possibility	Free text	
	How many contributions do you expect per day?	Choices	[less than 100 contributions, between 100 and 1000, more than 1000 contributions, I don't know]
	Could this data be interesting for any community?	Choices	[Yes, No, I don't know]
Yes (25)	To whom?	Free text	

**Table 1. List of questions for the DMP**

### **3.3 Data Management Workflow templates**

We have seen in section 3.2 how a user defines what is going to manage the data generated in its project. But, how is it going to do it? What are going to be the steps of the data lifecycle? In ACTION, a new tool is going to be deployed (based on templates) to facilitate this process. For this, the following elements will be necessary:



Lifecycle-aware citizen science templates

- A declarative format language to allow us to specify workflows. The Common Workflow Language<sup>8</sup> is our best candidate because it is widely used (to describe workflows) and many tools support it to define and execute workflows.
- Design a generic workflow to manage data in CS projects (see Fig.2)
- Similar to the DMP situation, CWL is a language that is not easy for non-experts to use. So, as with the DMPs, the solution is to create a web form to configure some parameters of the workflow.

In deliverable - D4.1 Conceptual Architecture - the steps of the participatory science lifecycle on a CS project were analyzed, generating a figure as a result. Next figure shows the part related with the research implementation, which is the important part in this section.

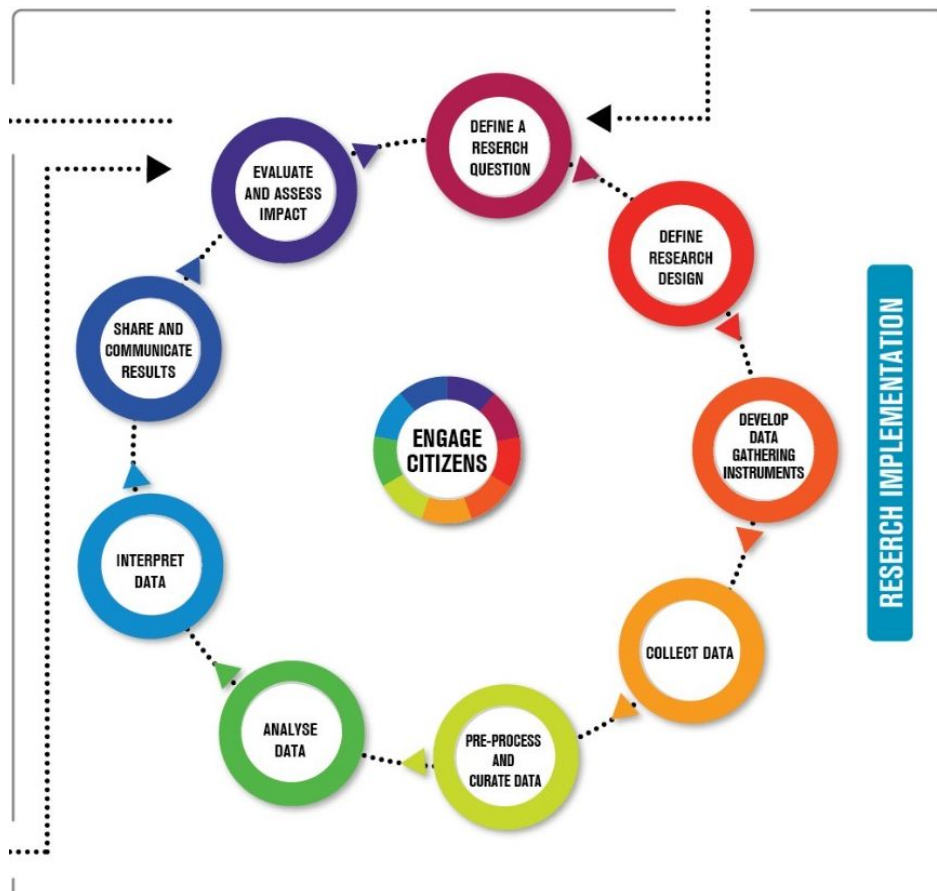


Fig 4. Participatory Science Lifecycle (data section)

This wheel represents the well known steps of the research process. Many of the steps involve data management, from the collection to their sharing and publication. As has been commented previously, the objective of this designed tool is facilitate these steps to the creators of CS projects.

<sup>8</sup> <https://www.commonwl.org/>





For that reason we are going to deploy a platform to execute simple workflows adapted to CS projects. The next figure depicts this architecture. First, a webform is used to configure the parameters needed for the workflow, similar to the DMP tool. This information will be complemented with the information provided by the DMP. Thus, in this case, the template will be a skeleton to generate the CWL file. This file will be executed periodically by an executor, generating a dataset and publishing it in a repository.

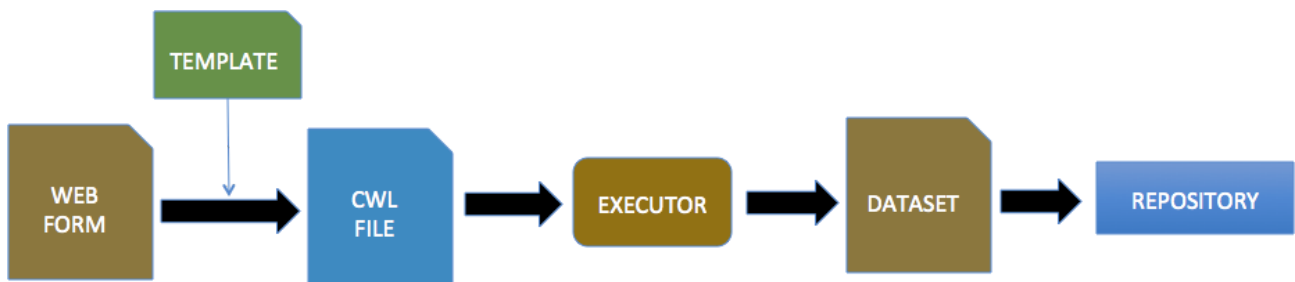


Fig 5. Generic workflow

Figure 3 depicts in more detail the steps that compose our generic workflow, which can be configured with different parameters (webform + DMP). As a note, the steps described in the figure are optional, depending on the user's execution. These steps are:

- **Harvester** module is in charge of generating a raw dataset based on the data present on different data sources. As can be seen in the figure, ACTION is going to develop some handlers (templates) to access different repositories or platforms. In our case, Epicollect (see deliverable *D2.1a - Tutorial to identify the spectra of common street lamps*), our internal databases or our data portal (available at M15). Users will be able to configure the period of data analyzed, etc .. Note that this information is provided in the DMP document.
- **Filter** module is in charge to filter the raw dataset in two criteria: based on a value of a specific column, or removing columns.
- **Divergences** module specifies what must be done with the fields that have divergences such as empty fields or problems with the format of the value.
- **Transformations module**: In this step, the user can decide to transform some data in order to prevent ethical or legal issues. In this case, we provide the system with two functions to anonymize personal data and obfuscate the location of the observations like in the case of publishing the location of protected species.
- **Publication module**: In this final step, the user decides where the platform has to publish the dataset generated. They can configure some fields such as the version, associated tags, license, etc ... Note that this information is provided by the DMP.

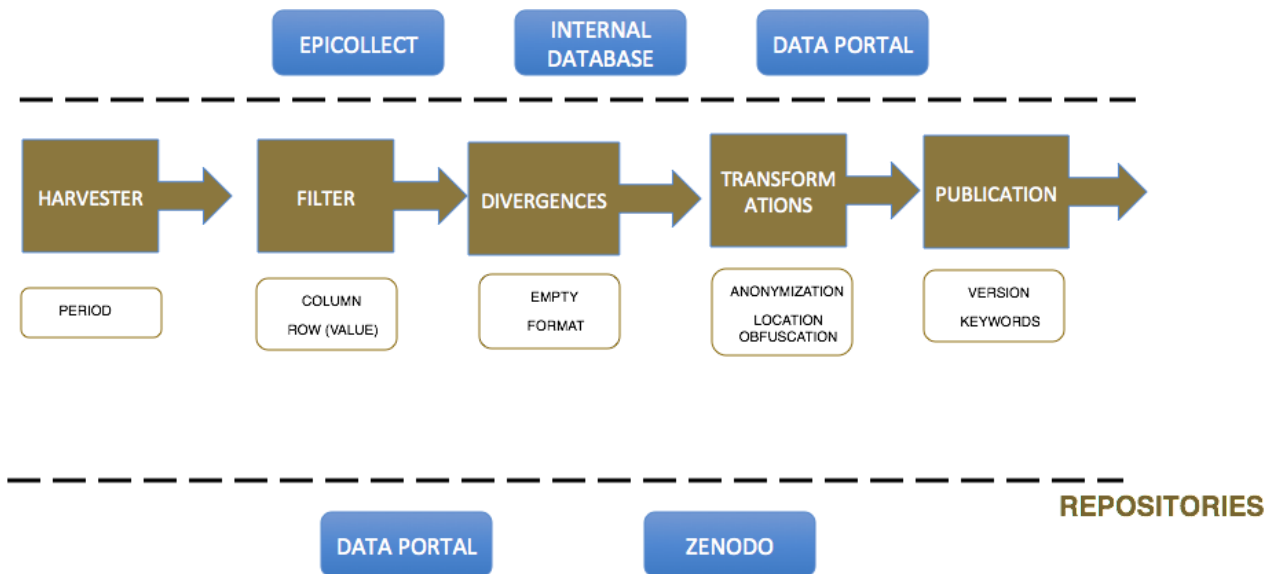


Fig 6. Sequence of steps to publish data

### 3.4 Presentation template

We have shown in previous sections some templates to initializing projects in Zooniverse, generate DMPs and define executables workflows. In Fig 4 we have observed that one of the steps is to communicate the results of the project. This template pretends to cover this part, creating an HTML webpage to show all the information related with the project, as well as the links with the results generated in the project. You can find an example here: <https://actionprojecteu.github.io/test-pilot-template/>. Github provides us with hosting for the webpages as well as a version control repository. Plus, the repo can be associated with the CS project’s users, which allows them to interact with the webpage.

This website will be filled in with the information provided in the rest of templates. The source code of the template can be found here: <https://github.com/actionprojecteu/actionprojecteu.github.io>

The structure of the website has the following structure

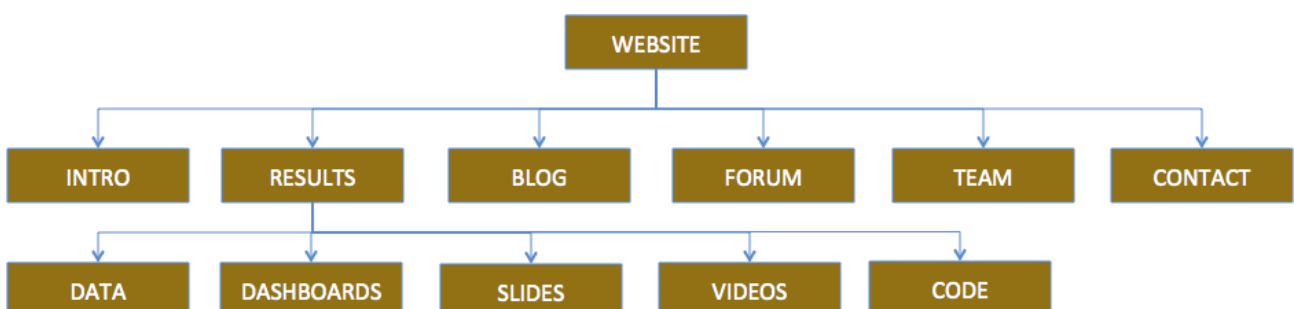


Fig 7. Structure of the website



Lifecycle-aware citizen science templates

Results section will be generated automatically based on the information provided by the users, or for the outputs generated in the project. At the end of the projects, results will be modeled as a Research Object<sup>9</sup> (*D4.4b - Research object catalogue v2 - M36*).

To host the website, we are using Github Pages.

---

<sup>9</sup> <http://www.researchobject.org/>

## 4 DMP creator tool

The objective of this tool is to facilitate the creation of DMPs for citizen science projects based on the information stored in ACTION and an interactive questionnaire filled in by users.

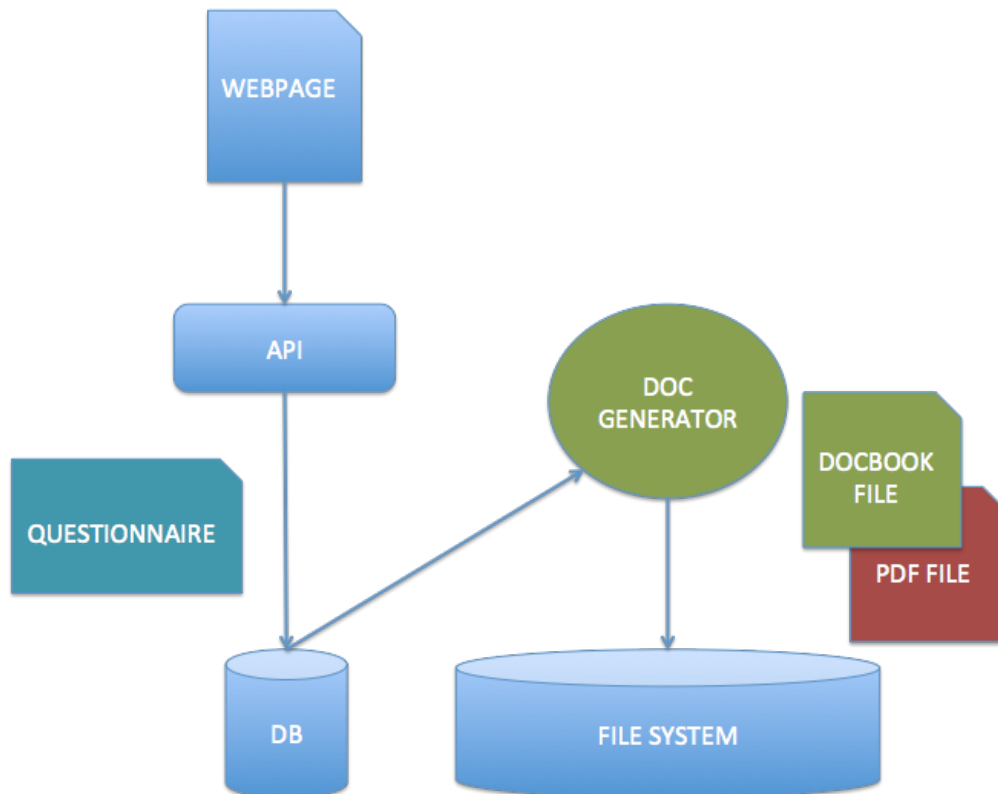
In order to develop this tool some requirements were defined in previous meetings with partners.

- REQ1: Only logged-in users can create data management plans.
- REQ2: The tool must be online and be responsive, it means, can be adapted to the size of the device
- REQ3: A user only can access to the DMPs generated by him
- REQ4: A DMP can be edited by its owner
- REQ5: The tool can generate a PDF document based on the information present in the tool.
- REQ6: The tool can generate a DOCX document based on the information present in the tool

For this first version of the tool, we have developed the following use cases:

- UC1: The user logs into the system.
- UC2: The user can answer to the questions present in a questionnaire
- UC3: The user can modify the answers of the questionnaire
- UC4: The user can generate a DMP document (PDF and DOCX) based on the answers to the questionnaire.

A typical web application architecture based on a frontend and a backend has been followed to deploy the tool. The following picture depicts the different components of the system



**Fig 8. Architecture of the DMP creator tool**

There are three main components:

- **Webpage**<sup>10</sup> (frontend) is built in Reactjs. Reactjs was selected because it provides faster rendering, easy maintenance and reduction of development times by its *components* approach. This webpage is a one page application to manage the DMPs of a user from the answer to a questionnaire (described in section 3.2).
- An **API**<sup>11</sup> (backend) built in Flask, the most popular web framework of Python. This API provides a set of methods to authenticate users, to create and generate DMPs.
- **DOC GENERATOR**<sup>12</sup> (in backend) built in Python. This module is in charge of generating a DocBook document (XML format) of the DMP, and later a PDF and DOCX document based on this DocBook. It has been designed as a background task, separated for the API, because generating the docs is a heavy operation that can block the API for other users.




The use of the tool is quite simple. After the user logs into the system, a window with a list of DMPs generated by the user will be shown. The user will be able to modify any of these, answer the questions again, or generate new versions of the documents clicking in the action buttons (see Fig 8). Also, a new questionnaire can be answered to create a new DMP (see Fig 9).

<sup>10</sup> <https://github.com/actionprojecteu/dmptool>

<sup>11</sup> <https://github.com/actionprojecteu/dmptool-api>

<sup>12</sup> <https://github.com/actionprojecteu/dmptool-generator>

### List of Data Management Plan

Name	Price
<b>test1</b> test2221ok	ACTION BUTTONS 
<b>test2</b> No description	
<b>test3</b> sdsds	

**Fig 9. Screen with the list of DMPs of the user**

test1 x

**What is the purpose of the data collected / generated?**

collection of elements 1

**Could you provide us with a description of the data you are going to generate?**

test2221ok

**Will you re-use any existing data?**

Yes  No

**What is the estimated size of each citizen contribution (in MB)?**

**How many contributions do you expect per day?**

**Could be this data interesting for any community?**

Yes  No

**Fig 10. Online questionnaire**

## 5 CONCLUSIONS AND NEXT STEPS

In this document, we have seen how the external platforms we plan to use do not completely cover most of the steps of our lifecycle of our projects. We identify in our pilots the necessity of focusing on data elements, so we are going to develop tools/templates in that direction. Plus, we discover that we have to put an extra effort in interconnecting the different tools analyzed through their outputs, which do not only include data, but also other elements such as presentations, papers, etc...

Regarding the DMP tool creator (in beta version) will be tested by the new projects in the next workshop that the project is going to celebrate in Berlin from 17 - 19 February.

The following table describes the different templates designed in this project and the deliverable version in which are going to be implemented.

The templates will be able to suffer modifications in later versions due to the interaction with the new pilots.

Category	Functionality	Version
Initialization	Create a project in Zooniverse	2
DMP	Web questionnaire (designed in chapter 3.2)	1
DMP	Generation of a DMP based on the questionnaire and the template	1
Workflow	Webform to generate a CWL file based on a template	2
Workflow	Execution of a workflow based on a CWL file	2
Workflow	Tool to manage workflows (creation, configuration, etc ...)	2
Workflow	Templates to step Harvest (epicollect)	1
Workflow	Templates to step Harvest (internal databases & data portal)	2
Workflow	Templates to configure steps Filter + Divergences + Transformation + Publication)	2
Workflow	Creation of tasks in Zooniverse	1
Web	Creation of a website	1
Web	Publication of results (based on research object catalogue)	2



Web	Tool to manage the website (introduction texts, forum, etc ...)	2
-----	---	---

**Table 2. Future steps**

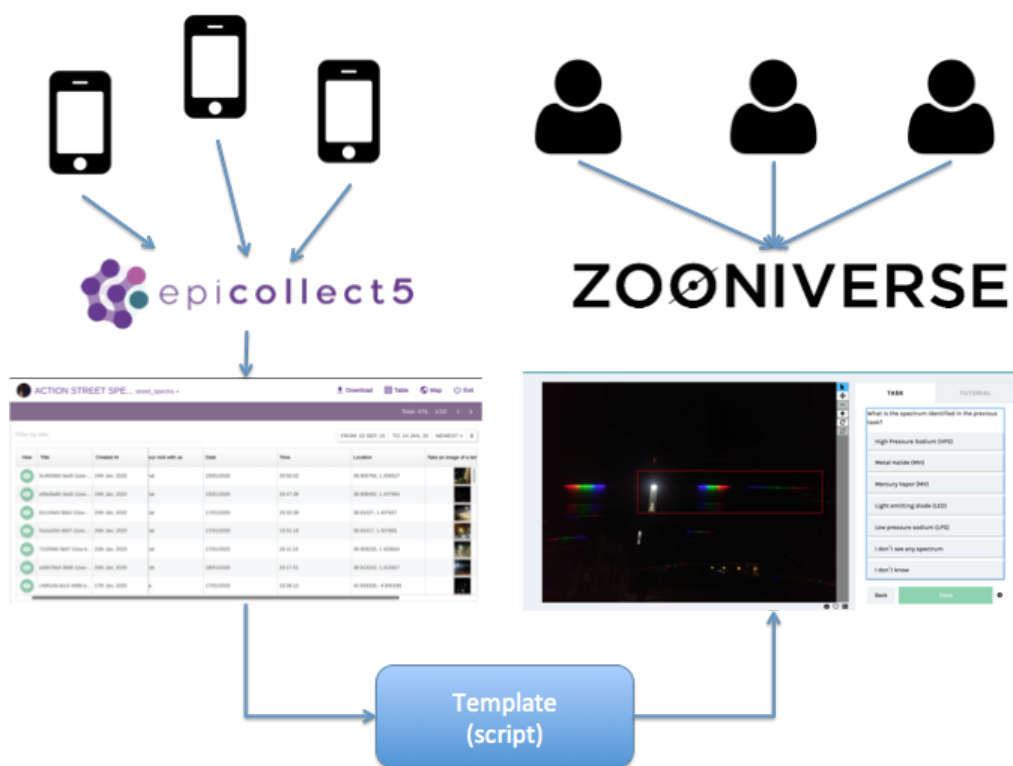
## **ANNEX I: Mini tutorial to create a CS project using Epicollect and Zooniverse**

## Lifecycle-aware citizen science templates

In this guideline, we are going to show how to create a system in ACTION to collect and process data using two external tools Epicollect5<sup>13</sup> and Zooniverse. To illustrate this process, we will use the example of street spectra, one of our pilots.

The StreetSpectra project wishes to turn the smartphones into scientific instruments to analyze lamps colors and their spectra. More information about this project, together with an extensive tutorial, can be found in D2.1 - Tutorial of street spectra project.

The next picture depicts the architecture of the system.



In this pilot, we have identified two use cases: i) volunteers will take the spectrum from a particular lamppost using the camera of its mobile phone and ii) volunteers will classify the spectra taken in the previous use case.

To connect both systems, a script has been developed to export the images (and metadata) taken with Epicollect to the Zooniverse platform, adding them as tasks in the Subject Set section. Note that this script has been deployed for this first version, in later versions, it will be part of the Data Management executor described in Section 3.3.

First, we are going to describe how to create a project in Epicollect to locate and to store spectrum from lampposts. Epicollect is a platform to allow users to create and easily design web forms to collect data.

<sup>13</sup> <https://five.epicollect.net/>





### Steps

1. Identify the data you want to collect. In our case we have identify the following data needed for this project (datetime, location, image, additional observations)
2. Log in the platform (<https://five.epicollect.net/login>). Epicollect uses google authentication as a login authentication system. For that reason, you will need a google account to create projects.
3. Once you enter into the platform, you will see a create project option in the menu. You will have to provide some extra information about your projects such as project name, small description and a name for the form. The name of your project must be unique in the system because it will be used as an identifier.

epicollect5 Hi, Esteban My Projects **Create Project** Find Project Logout

New Project Import Project

**Project name**  
i.e. 'My Survey 2015'  
Max 50 chars

**Small description** (A long description can be added later)  
what your project is all about...  
Max 100 chars

**Form name**  
i.e. Census  
Max 50 chars

**Access**  
 Private  Public

CREATE

4. Once the project is created, you can start to build your form. For that, select the option FORM BUILDER.

epicollect5 Hi, Esteban My Projects Create Project Find Project Logout

**ACTION STREET SPECTRA**  
Project homepage: <https://five.epicollect.net/project/action-street-spectra>

Dashboard

- Details
- Form Builder**
- Manage Users
- Mapping Data
- Clone
- Manage Entries
- Developers <sup>Beta</sup>
- API
- Apps

Project details

The light emitted by street lamps could be analysed using smartphone cameras, and other simple and i

Created on Monday 02 Sep 2019, 11:17

No description yet

Settings

**Access** PRIVATE PUBLIC

**Status** ACTIVE TRASH LOCK

**Visibility** LISTED HIDDEN

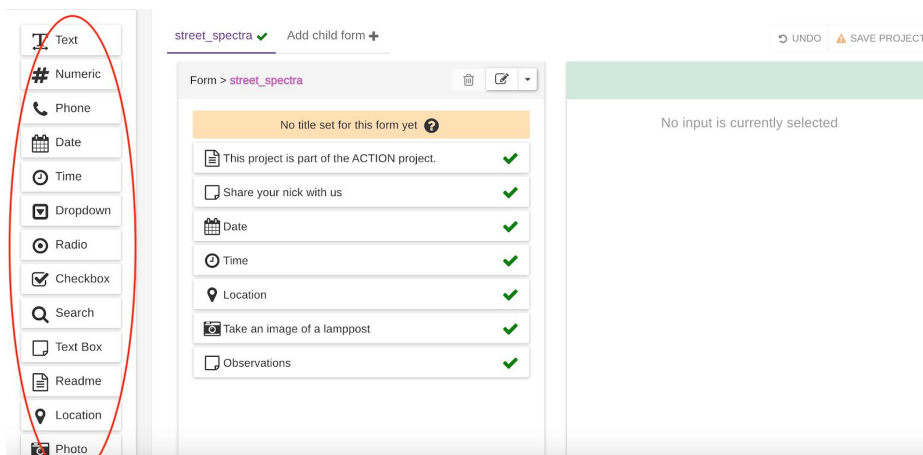
**Category** Science

Found a bug? Have a question? Post it on our community page!

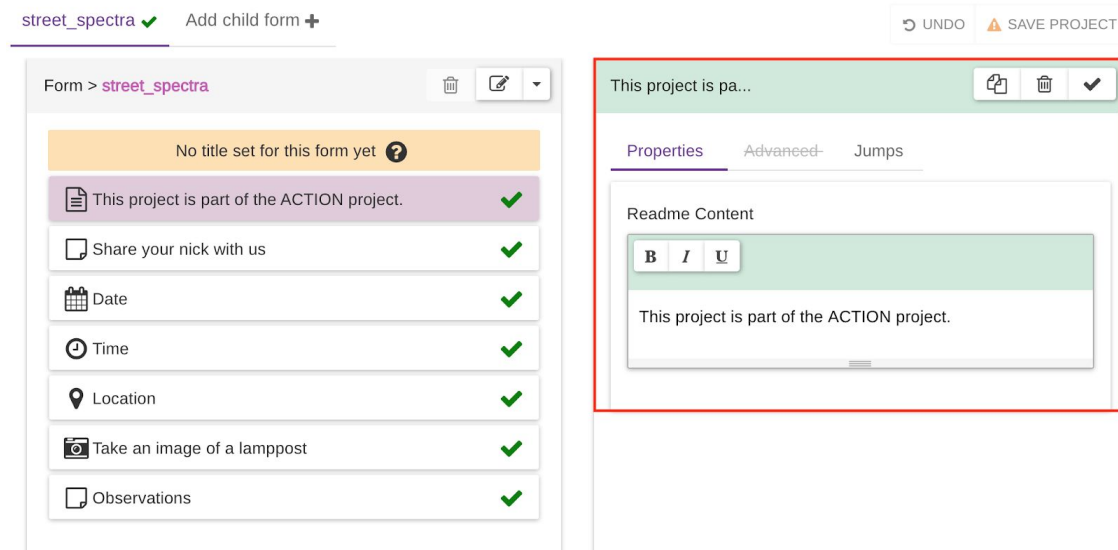
Do you like Epicollect5? Please consider rating us on the Play Store or App Store , thanks! 🍌



5. A set of fields will appear in the left side menu. Most of them are the typical fields in web forms such as Text, radio, dropdown and checkbox fields. But there are other interesting components to get location, current date or upload photos. Each component in the form will be shown to the volunteer as a new window.



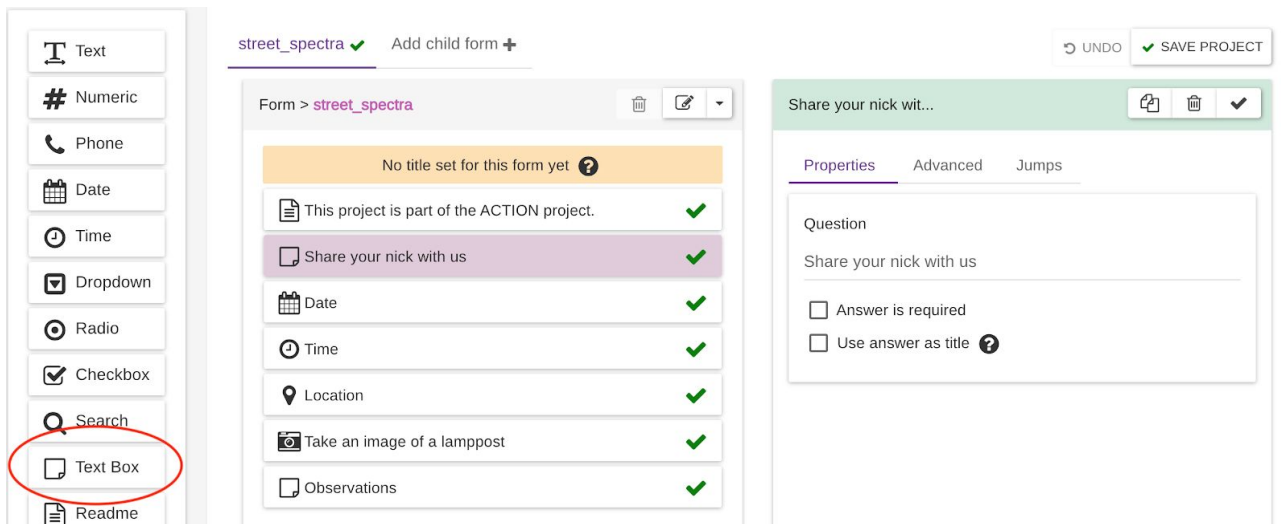
6. To add a field in the form, you have to drag and drop each component you need to the form (in the central part of the page) and configure it in the right side area.
7. The first component is a text area that will show a welcome message to the volunteer.



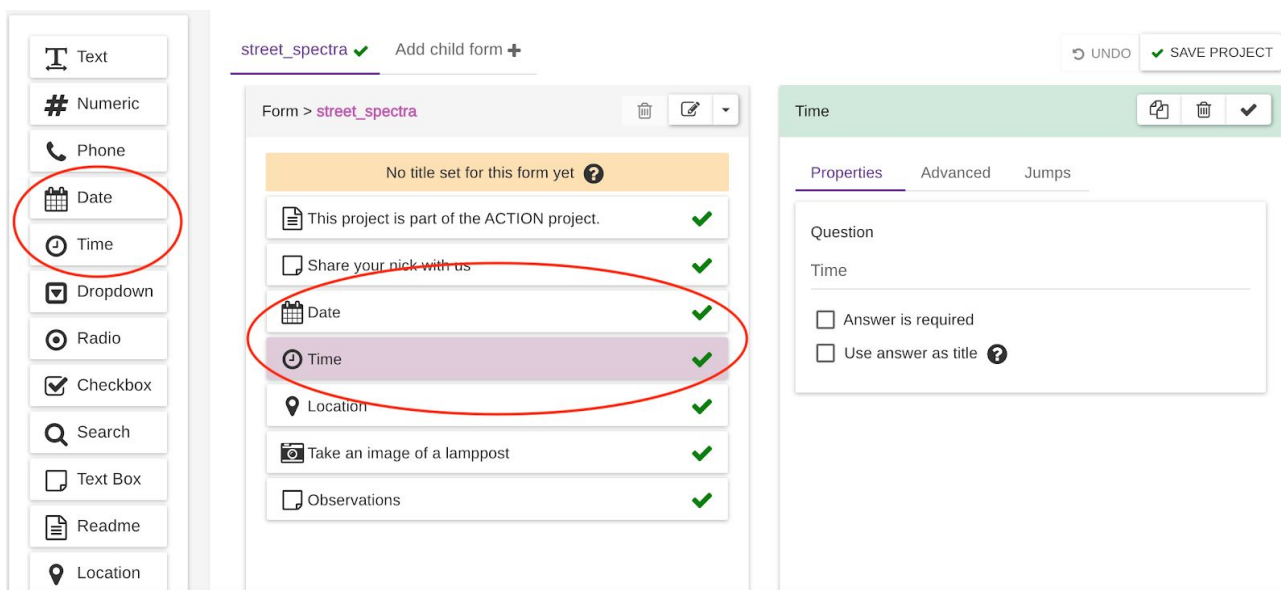
8. We need our application to get the nick of the volunteer. For that, you have to drag and drop the Text Box component, analog to the input field in web forms. To configure this component, we have to write the question that the application will do to the volunteers. In the advanced option, we can define a regular expression to validate the answer although it is not going to be used in our example.



## Lifecycle-aware citizen science templates



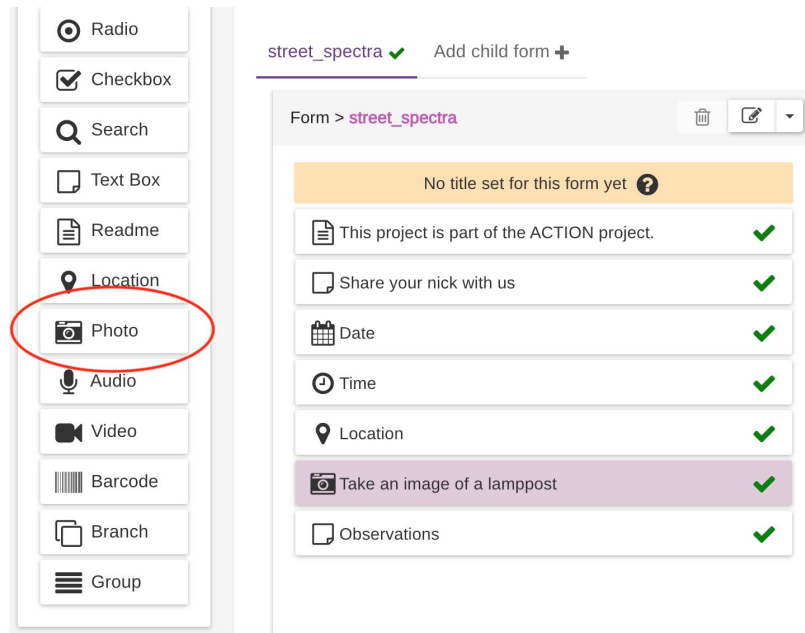
9. Following the data identified in the first step, the project needs the datetime of the observation. We will use two special fields: date and time. In the **Advanced** option we can define the date and time format. One interesting characteristic of these fields is that they can be automatically filled in with the current date and time when they are shown to the volunteer.



10. The next information we need is the location of the lamppost. For that, we will use the component Location. Similar to the date field, volunteers will not have to introduce the location, it can be taken automatically from the mobile.
11. Next important data is an image of the lamppost spectrum. We will select the image component. This component allows users to take images from the mobile camera or from the user's gallery. This image will be uploaded to the Epicollect servers.



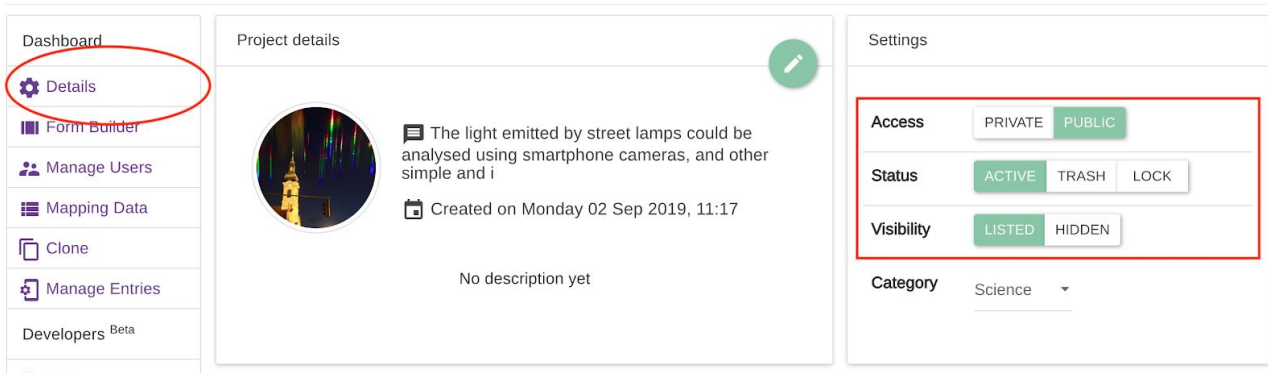
## Lifecycle-aware citizen science templates



12. Lastly, we will use a Text Box component to allow volunteers to write some relevant additional observations about the spectrum taken.
13. Once the form is complete, it can be saved by clicking on the *Save Project* button.
14. Finally, we need to publish the project so it can be listed and used with the application of epicollect. Go to the home page and configure these options: Access -> Public, Status -> Active and Visibility -> Hidden

### ACTION STREET SPECTRA

Project homepage: <https://five.epicollect.net/project/action-street-spectra>



The entire example can be found here: <https://five.epicollect.net/project/action-street-spectra>

All the collected information is openly available<sup>14</sup> to all users and can be downloaded in CSV or JSON format. Also, can be consulted in table format as can be seen in the following figure.

<sup>14</sup> <https://five.epicollect.net/project/action-street-spectra/data>



ACTION STREET SPE... street\_spectra

Download Table Map Exit

Total: 436, 1/9

Filter by title

FROM: 02 SEP, 19 TO: 30 DEC, 19 NEWEST X

View	Title	Created At	Time	Location	Take an image of a lamppost	Observations
	5cdc9541-9039-4461-...	30th Dec, 2019	20:08:29	40.616933, 20.78072		Korcë, Albania
	0ada66cd-144a-4f56-...	29th Dec, 2019	18:40:47	40.357391, -3.783624		
	6397bbff-5355-4052-9...	28th Dec, 2019	21:03:20	39.484582, -0.359485		LED muy blancos tipo mazorca.
	3b2aa76d-5381-4607-...	27th Dec, 2019	16:54:29	41.329733, 19.824692		Tirana, Albania
	60073dbf-bd09-4195-...	26th Dec, 2019	20:22:29	40.13082, -5.460953		
	1938f289-27ea-4009-...	25th Dec, 2019	19:58:37	40.354852, -3.813737		
	1c689b0e-2967-45a0-...	25th Dec, 2019	19:55:45	40.354557, -3.815465		
	6c4ee5dc-75ef-48a3-...	25th Dec, 2019	19:48:53	40.353809, -3.808964		

Fig XXX. StreetSpectra application in Epicollect

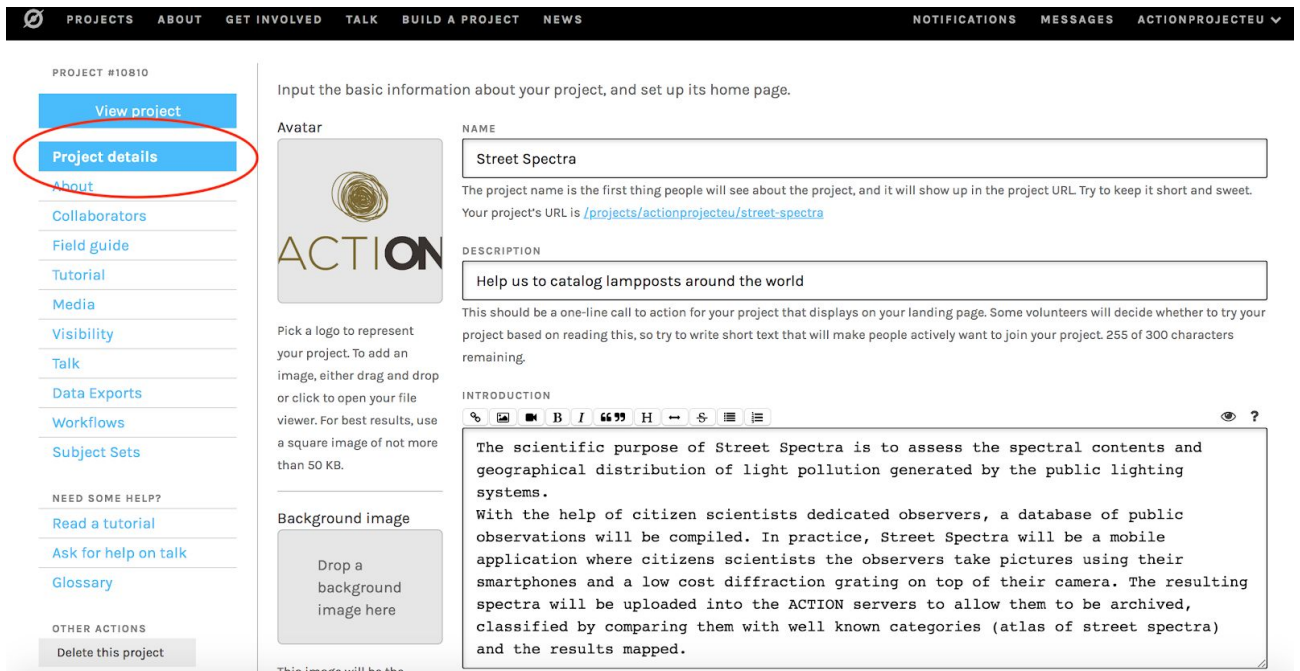
If you remember the second use case, the objective of our project in Zooniverse is to classify the images collected by users in epicollect5.

### Steps

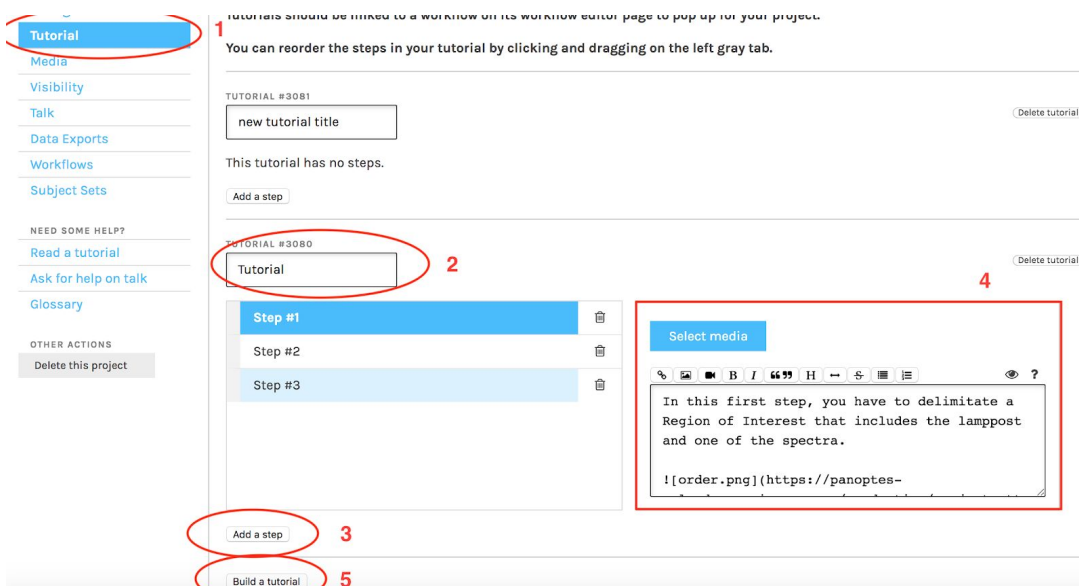
1. You have to sign up on the Zooniverse platform. In this case, the platform doesn't allow the use of the google authentication system.
2. Sign in the platform with the credentials used in the registration process.
3. Once you are in the platform, you can start to work in a project clicking in the BUILD A PROJECT option of the menu



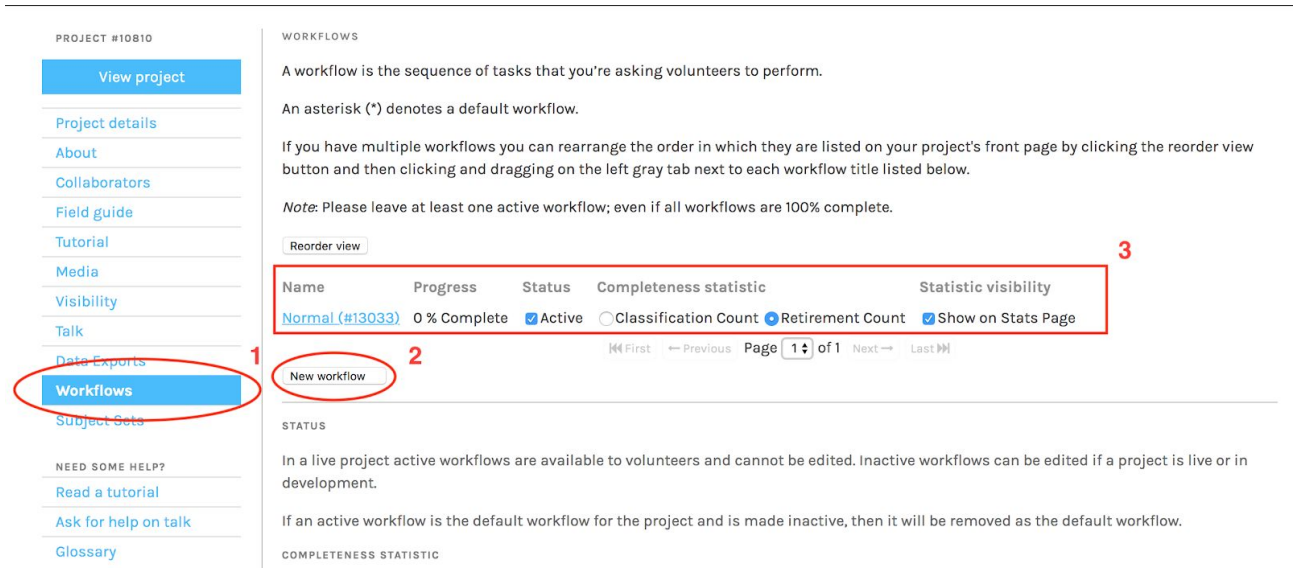
4. In the new window, you will have to set up the details of your project. It includes name, description, introduction and tags to make your project findable in the zooniverse platform.



- After that, we have to build a small tutorial to help volunteers to classify the images collected by Epicollect. First click in the *Tutorial* option (1) and write a name for your tutorial (2). You can add steps in the tutorial clicking in the Add a step button (3). Each step is composed by a text with images that can be configured in an online editor (4). After you have added all the steps, you can build the tutorial (5). The tutorial will be associated to an specific workflow.



- Once the tutorial is completed, we start to define the workflow. In our case, we are going to create only one workflow. You will have to click in New Workflow button (2) and a new entry will appear in the workflow list component (3)



PROJECT #10810

WORKFLOWS

A workflow is the sequence of tasks that you're asking volunteers to perform.

An asterisk (\*) denotes a default workflow.

If you have multiple workflows you can rearrange the order in which they are listed on your project's front page by clicking the reorder view button and then clicking and dragging on the left gray tab next to each workflow title listed below.

Note: Please leave at least one active workflow; even if all workflows are 100% complete.

Reorder view

Name	Progress	Status	Completeness statistic	Statistic visibility
Normal (#13033)	0 % Complete	Active	Classification Count	Retirement Count
				Show on Stats Page

New workflow

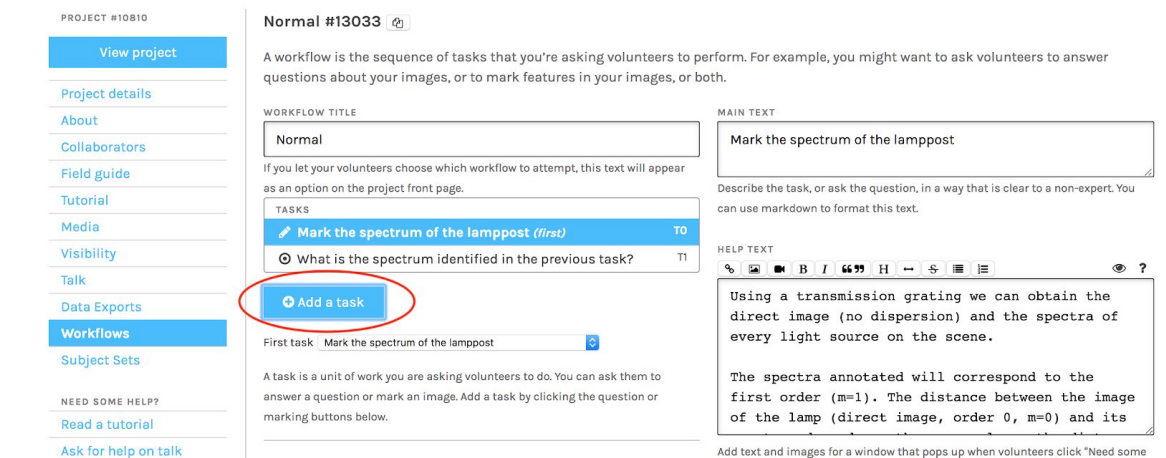
STATUS

In a live project active workflows are available to volunteers and cannot be edited. Inactive workflows can be edited if a project is live or in development.

If an active workflow is the default workflow for the project and is made inactive, then it will be removed as the default workflow.

COMPLETENESS STATISTIC

- In this one workflow, we will create one task for selecting a region of interest (which includes the lamppost and the spectrum). You will select the *drawing* template. Also, you have to select in the *Choices section* a rectangle as a type of marker.



PROJECT #10810

Normal #13033

A workflow is the sequence of tasks that you're asking volunteers to perform. For example, you might want to ask volunteers to answer questions about your images, or to mark features in your images, or both.

WORKFLOW TITLE

Normal

If you let your volunteers choose which workflow to attempt, this text will appear as an option on the project front page.

TASKS

Mark the spectrum of the lamppost (first) TO

What is the spectrum identified in the previous task? TI

Add a task

First task: Mark the spectrum of the lamppost

A task is a unit of work you are asking volunteers to do. You can ask them to answer a question or mark an image. Add a task by clicking the question or marking buttons below.

MAIN TEXT

Mark the spectrum of the lamppost

Describe the task, or ask the question, in a way that is clear to a non-expert. You can use markdown to format this text.

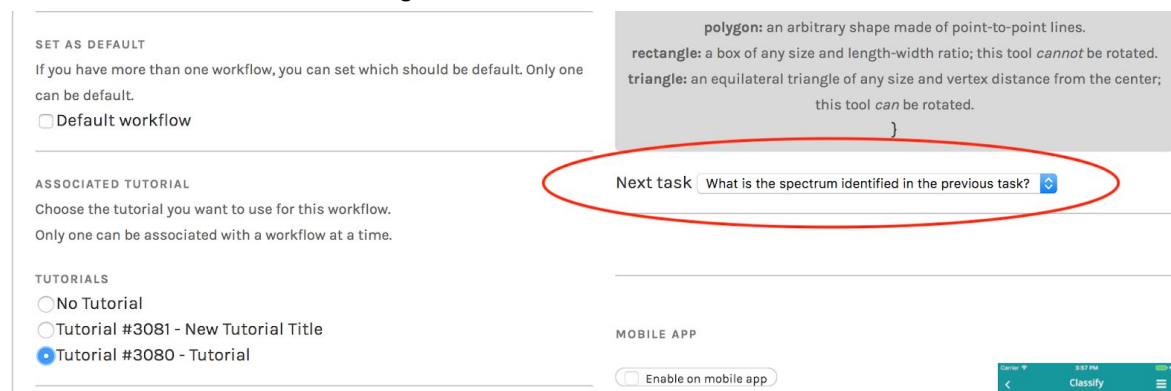
HELP TEXT

Using a transmission grating we can obtain the direct image (no dispersion) and the spectra of every light source on the scene.

The spectra annotated will correspond to the first order ( $m=1$ ). The distance between the image of the lamp (direct image, order 0,  $m=0$ ) and its

Add text and images for a window that pops up when volunteers click "Need some

- Now, we are going to create another task to classify the spectrum. We will have to repeat the process but in this case, we have to select the *question* template.
- Once the second task is created, we have to connect both tasks. So you have to come back to the first task and adding the second task as a Next task to be executed.



SET AS DEFAULT

If you have more than one workflow, you can set which should be default. Only one can be default.

Default workflow

ASSOCIATED TUTORIAL

Choose the tutorial you want to use for this workflow. Only one can be associated with a workflow at a time.

TUTORIALS

No Tutorial

Tutorial #3081 - New Tutorial Title

Tutorial #3080 - Tutorial

MOBILE APP

Enable on mobile app

Next task: What is the spectrum identified in the previous task?

Lifecycle-aware citizen science templates

10. Now, we have defined a workflow with the necessary tasks to classify the spectra. Next step is to feed the system with the images produced by the epicollect. This can be done manually in the subject sets section or automatically with the future tool of ACTION to manage workflows (see section 3.3)
11. The application is ready and can be published in the visibility section.

The application is available in:

<https://www.zooniverse.org/projects/actionprojecteu/street-spectra>

In this first version, we have created a python script to feed the system with images. In later version, this script will be integrated into the data management executor tool. As an example, the following picture depicts the schema of the process.

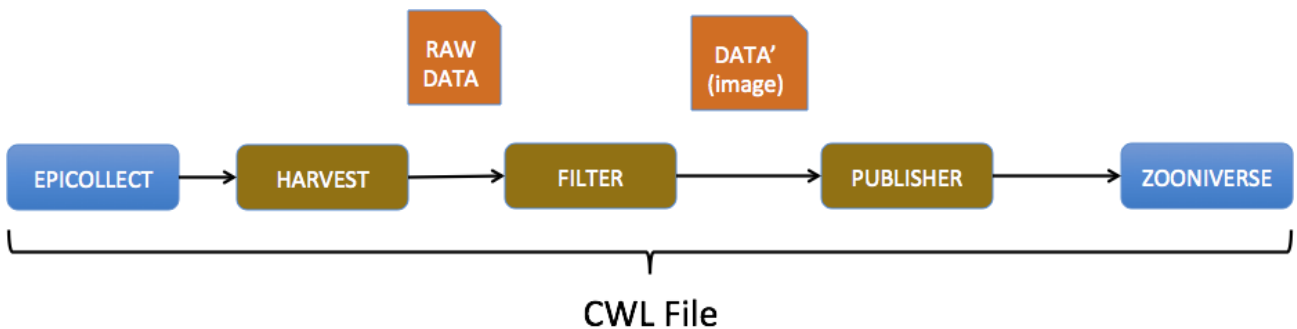


Fig XXX. StreetSpectra application in Epicollect



## **ANNEX II: Original set of questions in H2020 DMP template**



What is the purpose of the data collection/generation and its relation to the objectives of the project?
What types and formats of data will the project generate/collect?
Will you re-use any existing data and how?
What is the origin of the data?
What is the expected size of the data?
To whom might it be useful ('data utility')?
Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?
What naming conventions do you follow?
Will search keywords be provided that optimize possibilities for re-use?
Do you provide clear version numbers?
What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.
Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.
Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out. How will the data be made accessible (e.g. by deposition in a repository)?
What methods or software tools are needed to access the data?
Is documentation about the software needed to access the data included?
Is it possible to include the relevant software (e.g. in open source code)?
Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.
Have you explored appropriate arrangements with the identified repository?
If there are restrictions on use, how will access be provided?
Is there a need for a data access committee?
Are there well described conditions for access (i.e. a machine readable license)?
How will the identity of the person accessing the data be ascertained?
Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different available (open) software applications, and in particular facilitating



re-combinations with different datasets from different origins)?
What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?
In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?
How will the data be licensed to permit the widest re-use possible?
When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.
Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.
How long is it intended that the data remains re-usable?
Are data quality assurance processes described?
What are the costs for making data FAIR in your project?
How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).
Who will be responsible for data management in your project?
Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?
What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?
Is the data safely stored in certified repositories for long term preservation and curation?
Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).
Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?
Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?