Grant Agreement Number: 777483 / **Acronym:** ICEDIG
Call: H2020-INFRADEV-2017-1 / **Type of Action:** RIA
Start Date: 01 Jan 2018 / **Duration:** 27 months

REFERENCES:

Deliverable **D5.2** / [R] / [PU]
Work package **5** / Lead: **UTARTU**
Delivery date [M16]

ICEDIG.EU

Innovation and consolidation for large scale digitisation of natural heritage

Improving the detection of collection-based citizen science projects

Deliverable 5.2

Veljo Runnel[1]

Agnes Wijers[2]

1- University of Tartu Natural History Museum
2- Picturae

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

ICEDIG.EU

The deliverable comprises an analysis of existing citizen science project metadata standards (as PPRS-CORE) and of how the underlying collection-based biodiversity data and relevant digitisation efforts can be detected through this. Recommendations will be given for complementing the metadata with the elements for describing the digitisation efforts and the quality of biodiversity-related content.

Contents

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

1. Collection-based citizen science activities

Citizen science (CS) is generally understood as public, volunteer participation in scientific research[1,2]. Kullenberg and Kasperowski demonstrated that citizen science is composed of three main strands of research – in one strand biology, conservation and ecology, and the other two strands being geoinformatics and social science combined with epidemiology[3]. Natural history collection based research, however, does not emerge in their analysis. They also performed an analysis of term co-occurrences from which some very relevant terms, such as „collections“ or „transcription“ are missing. The more general term “crowdsourcing”, though, is one of the most frequent terms pairing with other CS-related terms. Out of 1935 articles only two were about collection transcription (indicated directly in the title).
To facilitate the current analysis, we will in the following sections describe the main types of collection-based citizen science that have direct impact on generating and mobilising of biodiversity data and the relative role of citizen scientists and professionals in this. Later, the aspect of data collection and data classification (or data interpretation) is taken into account separately, following Kosmala et al[4] who used this approach for ecological studies.

1.1. Private collections

Although most natural history collections are owned by museums (publishing 131 million specimen records to GBIF[5]), many collections are in private hands. The survey conducted under ICEDIG task 5.3 (Digitisation of small collections) revealed that the number of specimens owned by respondents in Europe is collectively between 8.8 and 32.6 million. This reflects the status of just part of European private collections, as only some countries participated and not all private collection owners responded to the questionnaire or were reached. As showed by ICEDIG survey, 84% of private collection owners share their collection data for scientific research. This indicates remarkable impact of this field of citizen science to the research.
The notion of private natural history collections is rarely touched in citizen science analysis, but it is justified to pay more attention to this facet in the future. There is the aspect of legal issues when collecting specimens from the nature, but if a collector is also a member of a society, good care is taken to educate members and collaborate with research institutions to gain scientific goals with collecting specimens.
The role of citizen scientists and professionals in the activities: citizen scientists are both collecting specimens and classifying data.

1.2. Collecting specimens (samples) directly for research institutions

Although specimen collection is sometimes mentioned as a part of citizen-science based monitoring programs and bioblitzes[6]**,** particular approaches are not well described in the literature, so the actual benefit of citizen science to specimen collection is difficult to estimate. Compared to the citizen-science contribution to biodiversity monitoring or the impact of private collections to scientific research this type of citizen science has probably only marginal role. T**he role of citizen scientists and professionals in the activities:** citizen

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

scientists are collecting specimens or samples, but professionals are classifying the data.

1.3. Crowdsourcing for label transcription

Crowdsourcing as a method is used for computing, medicine and other disciplines, including biodiversity research (natural history). It refers to a set of distributed production models that make an open call for contributions from a large, undefined number of people [7].
Crowdsourcing can be done on commercial platforms where the crowd is rewarded monetarily, like Amazon Mechanical Turk. In citizen science crowdsourcing is a voluntary activity and the reward is not financial. In the ICEDIG project the specific studies (MS26, D4.2[7]) analyse crowdsourcing platforms for specimen label transcription and we are not discussing it in details here unless the data standardisation is in question.
As crowdsourcing helps to enrich and mobilise collection data, which eventually will be published in open data repositories like GBIF, we will later take closer look at how (or if) the data from crowdsourcing platforms is transferred to data repositories.
Researchers can also use crowdsourcing for analysing specimens for a particular study. It that case citizen scientists are annotating certain features of specimens on the photos or classifying objects.
The **role of citizen scientists and professionals in the activities:** usually the material is collected by professionals and classification (interpretation) is done by citizen scientists.

As seen from the work of Kullenberg and Kasperowski, the main area where citizen science is contributing to collection-based research in the natural history domain, is label transcription via crowdsourcing. There are other types of crowdsourcing activities, based on collections, for example categorisation of bird specimens by their plumage[8], but these types of projects are rather exceptional.

2. Scenarios for linking the collection data and citizen science data

The main emphasis of the current study is to analyse data and metadata availability and compatibility when collection-based citizen science is involved. We distinguish two scenarios of how citizen science metadata and collection data can be linked in the process of data transfer. These two scenarios will be further analysed for metadata detection aspect when the data are shared in internet.

1) **Collection-based datasets are created or enriched at collection holding institution with the help of citizen scientists and published in a specific data aggregation portal** (e.g. gbif.org). Probably the most common case of this scenario is when a museum is using transcription platforms to engage citizen scientists and the specimen data are enriched during the process. The information about citizen science involvement (if present) is stored in dataset metadata. The collection-based dataset itself contains information about taxon occurrences and about additional information of the collecting event. Here it is obvious that dataset contains occurrence and/or sample data based on collection specimens but the information about citizen science involvement and implicated quality issues can be obscure or missing.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

2) Citizen science projects are managed and published (shared) via a portal and/or online workbench but datasets which are collected or produced in the projects are not aggregated in the portal (e.g. citsci.org). The information about topic, methods, responsible institutions etc. is stored in the project metadata. If a project is dealing with natural history collections and some data are collected and stored, datasets can be managed outside of portal. Ideally, metadata of the project is indicating also access to the collection-related data.

Next we will discuss the two workflows and the metadata storage during them in details.

2.1. NH collection datasets published to data aggregators

When creating or enriching a NH collection dataset with the help of volunteers, information on citizen science involvement can be stored for several reasons and effects. The main questions for our review are: how (if at all) this information is stored and how this information is transferred in data aggregation processes. These aspects about citizen science involvement play an important part in evaluating CS in data creation and enrichment. And in research in general, also allowing to evaluate quality assessment specific to citizen science.

1) **Recognizing and measuring citizen scientists' contribution**

Based on analysis of earlier studies, Kullenberg and Kasperowski claim that acknowledging the volunteers contributing to the research in scientific publications has been quite modest [3]. As the public acknowledgment is one of the key factors to motivate citizen scientists to sustain their participation[9], research communities should pay more attention to make it happen. Can we, with automated data flows from citizen science portals, discover CS participation in data collecting at all? If there is no sign in collection datasets that citizen scientists have contributed to data enrichment (for example label transcription), we also fail to give proper credit to the role of citizen science in society. It would be very likely that a researcher who is analysing collection datasets, fails to see CS contribution if it is not documented. Societal and political willingness to recognise and reinforce CS contribution will also diminish if the evidence of participation is lost in data transfer.
European Citizen Science Association has identified ten key principles which underline good practice in citizen science[10]. Recognition is one of them: "Citizen scientists are acknowledged in project results and publications." Australian Citizen Science Association has adopted those principles but is reviewing them annually. They have highlighted the aspect of recognition of citizen scientists in projects by stating: "Citizen scientists are suitably acknowledged by projects. This may include acknowledgement in project communications, result reporting and publications" [11].
We should mention that although recognition is one of motivational factors in citizen science, contributing to science and helping the environment or community can be even more crucial[12].

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

a) **Simple mentioning of the fact of citizen science contribution (R1)**
Citizen science participation can be notified in dataset description. This can help to assess overall extent of citizen science as method. Some studies have shown that there is a lack of case studies showing actual contribution of citizen scientists to research[13]. If the actual contribution is not reflected in metadata of datasets, it is easy to underestimate the role of citizen science in research.

b) **Recognizing and measuring the effort of citizen science contribution (R2)**
As a more advanced approach, citizen scientist's contribution to data collection or analysis can be measured. There are certainly different ways to do that. For example, if volunteer crowdsourcing was used for digitizing 80% of labels in dataset and 20% was transcribed by professionals, this can be pointed out in metadata. Also, the number of volunteers and their working time for a project can be measured and described. Quantifying the volunteer contribution can also help to calculate financial gains.

c) **Recognizing individual contribution (R3)**
As a courtesy, individual citizen scientists can be cited as contributors, much like expert help is transferred to research publication authorship. If there are tens or hundreds of contributors, which can be the case with crowdsourcing, making reference to those individuals can be challenging, but by, for example, using dynamic citation, this can be made possible[14].

2) Quality assessment and management (Q)

Although in current analysis we will limit our scope to NH collection related citizen science, it is important to mention that most of the research on citizen science data quality (DG) has been based on the monitoring and observation projects. Collection-based citizen science is covered infrequently**.** The challenges of citizen science in data collecting have been pinpointed by several researchers along with approaches to overcome them [4,15–19]. A researcher who is working with citizen science contributed data, can assess data quality and use complementary methods to repair biases or to filter out unfit data. As it is stated by GBIF representatives [20], detecting and correcting errors should ideally be managed by both – data aggregators and data providers. The latter could be a citizen science project manager at a research institution or a local data collector. If errors in datasets have been detected by data aggregator, provider should be notified and data amendment and republication should follow.

Two important aspects of data quality **– data quality assessment** (judging data fitness for use) and **data quality management** (action of improving the data quality) [21] – can be described in dataset metadata. When a citizen science contributed dataset has been shared to a data aggregator, data quality assessment can already be done by the dataset publisher or the institution that is responsible for the dataset compilation. If not, end users of the data need data quality measurements to assess data quality themselves. Veiga et al also propose data quality status reports. The data quality report describes the current status of quality of a dataset or a single record according to the perspectives of data users. Such report contains data quality measures, validations and amendments that enable data users to perform the data quality assessment. For the dataset, data quality assessment would not necessarily be connected to citizen science components.

Quality in crowdsourcing can be more generally defined as „the extent to which the provided outcome fulfils the requirements of the requester“ [22].

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Allahbakhsh et al[22] have identified the following **quality control approaches** in crowdsourcing which can be applied to volunteer transcription:

- **Effective task preparation**
  Easy to understand user interface of transcription platform to minimize task misinterpretation etc.
- **Worker selection**
  Competent and motivated transcribers have been recruited.
- **Expert review**
  A qualified expert will validate results.
- **Output agreement**
  E.g. several transcribers will have the same result transcribing one label or one text field.
- **Input agreement**
  Not applicable for HH collection crowdsourcing.
- **Ground truth**
  Results will be compared with existing comparable data (e.g. transcribed locality text will be compared with existing localities of region or country).
- **Majority consensus**
  The judgment of a majority of reviewers on the contribution's quality is accepted as its real quality.
- **Contributor evaluation**
  Contribution is assessed by the competency/quality of transcriber.
- **Real-time support**
  Real-time help for transcribers to increase quality (online helpdesk).
- **Workflow management**
  Task workflow is optimised to increase quality.

2.2. Citizen science portals for managing and publishing projects

There are many websites that have been created for collecting and displaying information about different citizen science projects or initiatives (called portals in current study). Some of them are managed only by website owners, others allow co-creation with the citizen science community. In the latter case, project managers can enter the project information themselves. Portals can also differ in the audience geography. There are national portals, global portals or national portals with global ambitions. National portals are often only presented in local language(s) and are difficult to grasp for foreigners. US, UK and Australian portals are in favour of having global reach, as English is an internationally used language.

In some countries, citizen science has wider official support and portals cover a range of research areas – biology, geology, literature, history, medicine etc. The German portal "Bürger Schaffen Wissen" [23] is an example of such overarching platform for dissemination of different initiatives in the country. In other countries every research branch can have their own citizen science portal which collect project information only on specific topics.

A citizen science website can be more specifically oriented on one aspect of activities, like

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

biodiversity observation portals are dedicated for monitoring species. In our work, we exclude these portals, as they are already declaring the data type they are dealing with. UK National Plant Monitoring Scheme[24] is an example of such webpage, dedicated to one initiative only. Also, we are mostly looking at individual websites, not just subsites of another organisation or initiative. There are some exceptions, when a subsite of official webpage can be the only citizen science "portal" in the country and implies a significant effort for disseminating citizen science projects.

Functionality of CS portals

The main functional features of CS portals which also support collection-based research are listed below:

- ***Registering*** *a project or activity.*
- ***Marketplace*** *- project managers are recruiting participants (transcriptors), citizen scientists are looking for suitable projects to participate.*
- ***Communication*** *with already engaged participants, giving instructions for participation, collecting feedback.*
- ***Dissemination*** *of project results – reports, blogs, videos, photos.*
- ***Collecting data*** *– portals can also offer the tools for data collection and management.*
- ***Publishing data*** *– data which are collected via portal can also be published in portal.*
- ***Rerouting*** *- if a portal is not hosting data collecting/management/publishing services itself, it can reroute participants to actual data site.*

We collected information on some global and national citizen science portals. In chapter 4.1 these portals are reviewed, keeping focus on how they present the information of projects with the help of search interface, filtering capabilities, categorisation of projects. Then we examine how search and filtering functionality of those portals can help us discover collection-based projects.

3. Data standards for natural history collection data and metadata and citizen science project metadata

In this chapter common standards for NH collection data and dataset metadata will be described and reviewed on how they could be used for describing citizen science content. Also the reverse will be done – how can CS project metadata reveal collection information if a project is about collection label transcription or collecting specimens?

3.1. NH collection-based datasets

Mainstream data standards for publishing primary biodiversity data (observation and sample-based taxon occurrences) are Darwin Core Standard (DwC) and Access to Biological Bollection Data (ABCD). DwC is used by data aggregator GBIF.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

One of the most widely used metadata standards for datasets in ecology domain is Ecological Metadata Languange (EML). GBIF has adopted it for its own dataset metadata, GBIF Metadata Profile (GMP), EML file is included in Darwin Core Archive package.

Citizen science related references in DwC and GMP.

We identified data fields in DwC and GMP which can be used for citizen science related references, which can further give us information about the role of CS in data collection or enrichment and also about quality of volunteer work (Table 1, 2). We used Darwin Core Archive Assistant[25] to review DwC data fields and their relevance to CS elements. Information on GMP metadata fields was retrieved from https://github.com/gbif/ipt/wiki/GMPHowToGuide

Table 1 Embedding citizen science information in DwC (collecting event/specimen level data)

Aspect of CS	Darwin Core term and values	Comments
Recognizing and describing the fact of citizen science contribution (R1)	eventRemarks	Can be used in broad sense for CS references.
Recognizing and quantifying the effort of citizen science contribution (R2)	samplingEffort	Only usable for CS, if the fact of CS involvement is given. Otherwise it is not possible to determine if this means professional or CS effort.
Recognizing individual contribution (R3)	identifiedBy	Can represent the name of the citizen scientist.
	recordedBy	Can represent the name of the citizen scientist.
Quality assessment (Q)	identificationVerificationStatus	HISPID vocabulary[26] can help to evaluate expertise
	samplingProtocol	May refer to citizen science methodology or transcription protocol etc.

Table 2 Embedding CS information in GMP (metadata level)

Aspect of CS	GMP term and values	Comments
Recognizing and describing the fact of citizen science contribution (R1)	keyword	Any keywords that describe citizen science contribution can be fused, e.g. “citizen science”, “volunteer” etc.
	designDescription	Volunteering in general can be mentioned here.
Recognizing and quantifying the effort of	-	

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

citizen science contribution (R2)		
Recognizing individual contribution (R3)	Personnel	A field which describes people who are involved. As it is just one data field not differentiating professionals or volunteers, it cannot be used for quantifying volunteer effort.
Quality assessment (Q)	methodStep	Steps of volunteer involvement for which quality control can be applied.
	qualityControl	Quality control applied for a specific step in volunteer involvement.

Conclusions

There are several options to refer citizen science methodology both on dataset level (DwC) and metadata level (GMP) but there is no clear guidance for using appropriate terminology.

3.2. Citizen science project metadata

Standardization efforts of CS project metadata are linked to major CS association in the world (see appendix). PPSR_Core is used as Citizen Science project metadata standard for large citizen science project catalogues as SciStarter, citizen science catalogue at Atlas of Living Australia (ALA), CitSci.org.

We reviewed PPSR_Core Interim Project Metadata Sharing Protocol and PPSR_Core dataset data model v.0.2.0, published as a part on report of Citizen Science Association in 2017[27] to identify the data fields that can be used for storing information about NH collections (Table 3). Current work on the ontology is available at COST Action 15212 WG5 Deliverable from 2018[28].

The metadata sharing protocol is used for sharing project information between catalogues, for example information of some global SciStarter projects is also exported to ACSA Project Finder page [29].

Table 3 PPSR_Core project metadata sharing protocol

PPSR_Core field name / database term	Description	Type	Implications related to NH collection discovery
BASIC PROJECT INFORMATION			
Aim / projectAim	Primary aim, goal, or objective.	Text	As this is not standardized text, collection content

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

			may or may not be revealed.
Description / projectDescription	Abstract or description of the project	Text	As this is not standardized text, collection content may or may not be revealed.
Tags / projectTags	Controlled vocabulary terms, supplied by the person who entered the project, to assist with search and filtering.	Vocabulary	As this vocabulary seems to be not yet standardized, collection content may not be revealed.
Keywords / dcatKeyword	Keywords (comma separated) which are indexed and aid in searching for and finding projects. Data Catalogue Vocabulary (DCAT)	Text	Relevant keyword can be inserted
Project topic / projectScienceType	The project topic or field of science.	Vocabulary	Vocabulary still incomplete.
Intended outcomes / projectIndendedOutcomes	A project's goals, or intended outcomes of participation.	Vocabulary	
SUPPLEMENTARY INFORMATION FOR CITIZEN SCIENCE			
***Participation tasks / projectTasks**	The full list of tasks that volunteers may do to contribute to a project	Vocabulary	Most suitably describes collection-related content with several matching terms (see below).

Vocabularies:

projectTags: freely assigned terms

projectScienceType: list available in 2017 does not yet comply with the guidelines for mutual exclusivity and non-ambiguity and is incomplete.

projectIntendedOutcomes: Civic and community benefit // Conservation // Individual learning // Programmatic // Research advancement.
In some cases, sub-terms are also used. This vocabulary is being developed by the DEVISE Project.

projectTasks:

Annotation // Audio or video recording // Classification or tagging // DIY hacking/making // Data analysis // Data entry // Download software for distributed computing projects // Finding entities // Geolocation // Identification // Learning // Measurement // Observation // Photography // Problem

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

solving // Sample analysis // Site selection and/or description // Specimen/sample collection // Transcription

Conclusions

Referring to NH collections in CS project metadata is not straightforward, vocabularies are still incomplete. Including collection references in the basic metadata fields is largely dependent on project managers who are filling in the information as free text.

4. Case studies - crowdsourced collection data at the aggregator portals and collection-based projects at the citizen science portals

4.1. Crowdsourced datasets

4.1.1. Crowdsourcing platforms

If a collection dataset is being enriched with the help of crowdsourcing, usually a transcription platform is used which is designed specifically for this purpose. In ICEDIG, two reports have been produced for showing functionality of platforms and their data exchange protocols. A comparison of six widely used transcription platforms (ICEDIG MS26) is available at https://dissco.teamwork.com/#files/5517073. The data exchange between collection management system and transcription platform is described in ICEDIG MS28, "Specification of data exchange format for transcription platforms" [30].

Next we will review some crowdsourcing platforms in regard of their capabilities of citizen science metadata storage and publishing. Reviewed platforms: DoeDat (Belgium), Les Herbonautes (France), DigiVol (Australia). In addition, the description of globally less known platform Vele Handen (Netherlands) is included as appendix. Results are provided in Table 4.
Transcription platform representatives where contacted by e-mail and following questions were asked:

Volunteer involvement in general

If volunteering as a method was used in data producing or enriching, even when partially, does dataset metadata explicitly state this? Related terms may be used like "citizen science", "public participation" etc.

Do you store this information locally?
Do you publish this information (GBIF or other aggregators)?

Measuring volunteer contribution

Do you describe the effort (working hours, number of specimens processed etc.) of volunteers? If data in dataset are produced or enriched by both professional experts (paid) and volunteers, effort

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

can be also expressed as ratio – number of professionals / volunteers in the team, working time of professionals/volunteers or as a subjective assessment by project manager.

Number of volunteering individuals

Do you give the number of volunteers who are involved in transcription project of a specific dataset (if so, do you also distinguish between transcribers, team leaders and other supportive volunteers)?

Identity of volunteers

Are the names and addresses or e-mails of volunteers stored?

Description of quality control measures

Are quality control measures described which help to assess dataset quality? (E.g. expert review, output agreement among transcribers etc.)
Remark: in some case citizen science quality control is not much different from standard quality control in research.

Results:

Table 4 Metadata recording and storing in local system of transcription platforms

Transcription platform	Volunteer involvement	Measure of volunteer contribution	Number of volunteering individuals	Identity (names, e-mails) of individuals (GDPR issues)	Description of quality control measures
DoeDat	Yes	No (but some information can be inferred from raw data)	No	No	No
Les Herbonautes	Yes	No	Yes, as mission "members"	No (participating in mission does not need registering name)	Not explicitly
DigiVol	Yes	Yes, but not publicly available	Yes	No	Yes
Vele Handen	*Information not available*				

Metadata transfer to data aggregator (GBIF etc) (is data, stored locally, transferred to GBIF?)

Transcription platform	Volunteer involvement	Extent of volunteer contribution	Number of volunteering individuals	Identity (names, e-mails) of	Description of quality

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

		from overall effort		individuals (GDPR issues)	control measures
DoeDat	Data are not transferred to aggregator yet				
Les Herbonautes	Data are not transferred to aggregator				
DigiVol	The use of DigiVol is attributed in metadata	No	No	No	No
Vele Handen	Data are not transferred to aggregator yet				

4.1.2. Crowdsourced datasets at the aggregators

As for the three aforementioned transcription platforms, at the moment, there are no data workflows from DoeDat to publish data at aggregators (via the CMS), neither for Vele Handen. Les Herbonautes does not have neither clear pathway for metadata transfer to GBIF.

Atlas of Living Australia publishes datasets which have been transcribed with DigiVol platform. One such dataset was identified and respectable metadata was reviewed as follows.

Atlas of living Australia datasets https://collections.ala.org.au/

When searching for datasets, available et ALA , some filters can be applied which help to search crowdsourced datasets:

- Crowd sourced
- Transcribing complete
- Validation complete
- Validation in-progress

Filtering of crowdsourced datasets results 501 datasets (22.4.2019)

When viewing a specific dataset, for example *Australian Museum Tag a Fish-1,* we can see some basic metadata fields. Field “Description” contains some elements which help us to recognise it as a citizen science enriched dataset (R1), but nothing more specific. Data field “Content” displays the same categories which were applied when filtering search results (including “crowd sourced”).

Data field “Citation” contains link to DigiVol platform which was used for crowdsourcing https://volunteer.ala.org.au/

Conclusions

At the moment, most transcription platforms do not have definitive pathways for transferring citizen science related metadata to data aggregators and data transfer is done manually. It leads to assumption that datasets at the aggregators may miss important attribution to citizen science and

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

this may go unnoticed to stakeholders or the actual involvement of citizen scientists may be underestimated.

4.2. Citizen science project portals

As explained earlier, citizen science portals can vary in their content and reach. Many of them have potential to host also collection-based projects. However, the effectiveness to use a portal for advertising the collection-based projects depends on the search or filtering functions of the portals and how the projects are categorised on the site. A volunteer who is looking for a suitable collection-based transcription project from the portal should be able to find it quickly and in straightforward fashion.

Next, we will examine some citizen science portals in regard of discoverability of collection-based projects. For the purposes of current study, a portal should comply to following conditions:

- Portal represents diversity of projects (outdoor, indoor, web-based, different topics etc)
- Projects in portal are not representing a singular institution
- Targeted for wider audiences

To collect information about widely used and referred citizen science portals, an inquiry was submitted to the mailing list of European Citizen Science Association. This resulted in only ten responses with multiple portal references each. Most of these references where on specialist portals (observation portals e.g. Spanish Natusfera or European alien species projects website) which did not comply to study definition of a portal. Additionally, a general google search was performed with search phrase “citizen science portal”. Wikipedia article “Citizen Science” references to citizen science portals [31] were taken into account.

There are only a few global CS portals which can be considered as well-known and widely used. Some of them are primarily as national portals, but have been expanded to include also projects from other countries. Zooniverse, Scistarter, CitSci and ACSA project finder were examined for current study.

The list of national portals is longer, we included in our review national portals of Germany (Bürger Schaffen Wissen) and France (OPEN) as both countries have active citizen science communities. Unfortunately, the UK does not have a general national citizen science project portal, but as their Environmental Observation Network (UKEOF) is considered a major Citizen science hub, the portal was also reviewed. UKEOF shares information for both citizen science and research projects (called “activities”).

4.2.1. List of reviewed portals

Name	Geography	Scope	Number of CS projects (05.2019)	User content [1]

[1] Access of registered users to add projects to the portal

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Zooniverse	Global	Wide range of projects	89	Yes
Scistarter	National, USA / global	Wide range of projects	1669	Yes
CitSci	Global	Citizen science portal	774	Yes
ACSA Project Finder	National, Australia / global	Mostly biodiversity observation projects	437	No
OPEN	National, France	Mostly biodiversity observation projects	146	No
Bürger Schaffen Wissen	National, Germany	Wide range of projects	114	No
UKEOF (Environmental Observation Framework)	National, UK	Biodiversity observation projects	156	No

List of reviewed portals with respective url

Zooniverse	https://www.zooniverse.org
Scistarter	https://scistarter.org
CitSci	https://citsci.org
ACSA Project Finder	https://biocollect.ala.org.au/acsa#
OPEN	https://www.open-sciences-participatives.org/home/
Bürger Schaffen Wissen	https://www.buergerschaffenwissen.de/projekte
UKEOF	https://catalogue.ukeof.org.uk/

Results

The availability of search, filtering and categorisation options for projects in each portal was examined.

Search/filtering engine

Zooniverse	Simple search by project name. Filtering by activeness – popularity, launch time etc.
Scistarter	Simple search by parameters "Tags" and "Topic". Results can be further filtered by parameter "online only". Advanced search/filtering by parameters “Phrase/tag”; “Location”; “Projects to do while” (predefined values), “Topic” (predefined values); “Only projects that...” (predefined values); “Age groups” (predefined values).

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

CitSci	Simple search (any text). Result sorting by name, description, start date and measurement number.
ACSA Project Finder	Filtering by eight pre-defined parameters. Additional filtering by geography (area select tool).
OPEN	Simple search by keyword. Filtering by five pre-defined parameters – "Region", "Topic", "Species group" and "Difficulty level".
Bürger Schaffen Wissen	Filtering by topic and locality of projects. There is also option to filter out projects for children and finished projects.
UKEOF	Simple search, some boolean operators allowed. Filtering by 6 parameters: Type of records (Activity, Programme etc), Environmental domain, Purpose of collection, Catogory (Citizen Science or all), ECV keyword, Lead organisation.

Categorisation of projects, categories related to NH collections

	Number of project categories	NH collection related categories
Zooniverse	11	Biology, Nature
Scistarter	28	Biology, Animals, Birds, Ecology and environment, Nature and Outdoors
CitSci	-	-
ACSA Project Finder	30	Biodiversity, Animals, Birds, Geology and soils, Insects and pollinators, Information and computing sciences, Marine and terrestrial, Nature and outdoors
OPEN	13	-
Bürger Schaffen Wissen	16	Animals, Plants
UKEOF	9 environmental domains	Not directly compatible

Test search by keywords

To examine if collection-based projects can be detected by search engine, specific keywords were used. As the search with keyword "collection" revealed projects which apparently were not about natural history collections, the list was manually cleaned and irrelevant projects were excluded. Cleaned result is the number of projects which are strictly concerning natural history collections. The same was applied for the search results with keyword "transcription".

Search keyword	Scistarter	OPEN	CitSci	UKEOF	Bürgen Schaffen Wissen	Zooniverse	ACSA project finder
"collection"	163	12	28	82	Nota	Nota	Nota

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

"collection", cleaned	20	1	0	82[2]			
"digitization"	8		0	2			
"digitisation"	2		0	2			
"transcription"	12		0	0			
"transcription", cleaned	8		0	0			
"museum"	50		1	62			
"crowdsourcing"	32		2 (not transcription projects)	0			
"annotation"	3		2	0			
All projects	1669	146	774		114	89	437

Conclusions

Citizen science portals cover hundreds of projects each, most prolific is SciStarter with nearly 1700 projects listed. SciStarter also allows users to add and manage their own content (project managers). This is crucial when advertising the project and reaching out to the stakeholders.

When looking at the way portals allow to find NH collection related projects, results are rather underwhelming. Three portals from seven do not offer the search by keyword at all. Six portals have some sort of filtering options, some being more complex, some very simple. Most useful filtering option is by the domain (or research subject) of project. The categories under this filtering option vary by portals, most common relevant categories being “Biodiversity” and “Biology”. For portals with smaller numbers of projects the filtered results can be inspected manually to find suitable projects, with larger portals this can be unwieldly. Some filtering options are not relevant to the cause, e.g. the popularity of project or region of the project.

Test search in portals revealed that a rather small number of projects are actually about NH collections. Keyword “digitization” gave a disappointing number - 8 projects out of 1669 - at SciStarter portal.

Zooniverse portal also acts as data collection portal but houses relatively small number of projects. Although it has limited options for project discovery, the site compensates it with visually striking introductions of projects, which makes it more likely that users will browse through the project introductions just out of curiosity.

[2] Recording schemes of various species groups refer also to the work with museum collections, but it is not clear if volunteer work involves transcription or sample collecting.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

5. Main findings and recommendations for DISSCO

5.1. Acknowledging, attributing and measuring citizen science in collection datasets.

Label transcription by crowdsourcing is one of the main inputs of citizen science to the collection-based research. Giving proper acknowledgment and fair assessment to citizen science methodology is crucial for developing this line of cooperation and also establishing stronger social support for natural history research.

Darwin Core and GMP data standards both allow to describe and measure some aspects of citizen science input to some extent, but it needs clearer guidance, including appropriate terms in the vocabularies. The developments in metadata standardisation by major citizen science associations in Europe, USA and Australia are helpful for building a suitable vocabulary.

In any development of data and metadata standards for natural history collections it is important to refer citizen science involvement in the creation and enrichment of datasets, to refer actual effort and quality control measures in the relevant citizen science projects. The latter will make volunteer involvement more overall transparent and can help policy makers to assess usefulness of citizen science methodology.

During the (meta)data transferral from one system to another it is important that at least a minimum amount of information about citizen science involvement is included and it can be detected in the final publication of dataset. One of the most common system transferrals pathways is from transcription platform to collection management system and from there to data aggregators.

Although private NH collections play an important role in building material evidence for taxonomy worldwide, it goes largely unnoticed as a form of citizen science. The situation needs more attention; reviewing all the relevant data standards and adding private collection support in that regard is recommended.

5.2. Collection-based projects in citizen science portals

Citizen science portals are important hubs for volunteers who are looking for suitable projects to participate; they also help to build feedback system for participants. Unfortunately, collection-based citizen science projects are not easy to find in major citizen science portals. Not all portals feature a comprehensive search engine and filtering options in portals are often rather general. Considering that transcription portals have high publicity value themselves, it is not so much of a problem and citizen scientists will likely find the transcription portals without citizen science portal. However, for fair presentation of natural history collection holding institution in citizen science portals, it is recommended to work closely together with CS associations, propose the best solutions for keywords used in search engines and vocabulary to describe collection-based projects.

Appendix 1. Development of Citizen Science metadata standards

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Globally there are three major and influential citizen science associations – **Citizen Science Association**[32] (CSA) in USA, **Australian Citizen Science Association**[33] (ACSA) and **European Citizen Science Association**[34] (ECSA). While CSA and ACSA also act as national associations, ECSA is an organization which solely unites citizen science activists and stakeholders from different countries and with different cultural backgrounds. In Europe there are also local citizen science associations or centers but usually they are also associated with ECSA as well.

ACSA was formally founded in 2016 although its founders started active work already in 2014. ACSA is funded by governmental grant and its headquarters are situated at the Australian Museum[35].

The formation of **CSA** goes back to 2012 and it is affiliated with Schoodic Institute[36]. Among several working groups CSA is hosting, is Data and Metadata Working Group. One of the top priorities for this working group has been advancing CS project metadata standard called PPSR-Core.

ECSA was launched during EU Green Week in 2013 and has 8 working groups, one of them being Projects, Data, Tools and Technology. ECSA is registered as a non-profit organization in Germany is strongly affiliated with the Natural History Museum of Berlin.

As of 2018, all major citizen science organizations have focused their effort on developing a single standard for citizen science project metadata – PPSR_Core (PPSR standing for Public Participation in Scientific Research) or **PPSR Common Conceptual Model** (new name adopted in June 2018). Initial PPSR_Core data sharing protocol was intended to help data exchange between SciStarter, CitSci.org, Atlas of Living Australia and Federal Catalog (USA).

Currently several working groups have been involved in developing PPSR Core, to ensure that data is interoperable and usable in larger research community. There is an initiative to form **Citizen Science Global Partnership**[37] which also advocates supporting development of citizen science data standards.

In 2015, the U.S. Citizen Science Association (CSA) founded the **International Data and Metadata Working Group** (IDMWG) to promote collaboration in citizen science through the development and/or improvement of international standards for citizen science data and metadata. CSA DMW also works in cooperation with standards developments organizations such as OGC, W3C, TDWG.

In Europe, **COST Action 15212 "Citizen Science to promote creativity, scientific literacy, and innovation throughout Europe"**[38] embodies a working group which also is working on data and metadata.

European Citizen Science Association also has a working group for data and metadata with overlapping membership of abovementioned working groups.

Appendix 2. Research papers on citizen science standards

Title	Year	url

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Digitization protocol for scoring reproductive phenology from herbarium specimens of seed plants	2018	DOI: 10.1002/aps3.1022
Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary)	2017	DOI: 10.3390/s17112478
Upside-Down GIS: The Future of Citizen Science and Community Participation	2016	DOI: 10.1080/00087041.2016.1243863
How to tackle the molecular species inventory for an industrialized nation-lessons from the first phase of the German Barcode of Life initiative GBOL (2012-2015)	2016	DOI: 10.1139/gen-2015-0185
Citizen science networks in natural history and the collective validation of biodiversity data	2016	DOI: 10.1111/cobi.12696
Using phenocams to monitor our changing Earth: toward a global phenocam network	2016	DOI: 10.1002/fee.1222
A conceptual approach to a citizens' observatory - supporting community-based environmental governance	2014	DOI: 10.1186/1476-069X-13-107
Assessing predictions of population viability analysis: Peregrine Falcon populations in California	2014	DOI: 10.1890/13-1323.1
Annotating biodiversity data via the Internet	2013	DOI: 10.12705/626.4
Buckets of Resistance: Standards and the Effectiveness of Citizen Science	2010	DOI: 10.1177/0162243909337121
Standardized Information Models to Optimize Exchange, Reusability and Comparability of Citizen Science Data. A Specialization Approach	2017	https://portal.opengeospatial.org/files/?artifact_id=70328
Focal Plant Observations as a Standardised Method for Pollinator Monitoring: Opportunities and Limitations for Mass Participation Citizen Science	2016	https://doi.org/10.1371/journal.pone.0150794
A novel citizen science approach for large-scale standardised monitoring of bat activity and distribution, evaluated in eastern England	2015	DOI: 10.1016/j.biocon.2015.06.009

Appendix 3. Vele Handen, a case study

1. About *Vele Handen*

To facilitate crowdsourcing of heritage collections, Picturae started in 2011 velehanden.nl: a crowdsourcing website for the heritage sector. On this website every heritage institution with a collection can offer a project for access to the 'crowd', or the general public.
For heritage institutions, crowdsourcing is a wonderful solution to unlock collections. Many archives or museums have extensive collections that do not qualify for digitization and indexing, because it is too much work for employees and volunteers. Making use of the general public via crowdsourcing, this suddenly becomes possible. After all, many hands make light work.

If the project has been set up at *Vele Handen*, the volunteers can start working from their own computer at a time that suits them. The requested work varies from indexing to image selection. What is special is that while everyone can participate, the quality of the delivered work is unprecedented. The passion with which the volunteers work on *Vele Handen* should not be underestimated.

Participating institutions can choose to what extent they want to be involved in crowdsourcing. Picturae can arrange everything from digitization to completion of the unlocked project. Nevertheless, most institutions choose to actively contribute to the project. The interaction with the volunteers is found to be very nice and valuable. If the work is completed and the project is completed, the project will be retrieved from *Vele Handen* and the unlocked project will be

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

returned. To this end, the institution can decide to integrate the collection within its own website or to present it on a new website.

1.2 Tools within Vele Handen

Indexing
Using pre-formed forms, the volunteer collects data, for example 'name', 'place' and 'date of birth' on the basis of the scanned source.

Linking
Volunteers ensure that different data are linked to each other. For example, a personal name is linked to the correct register.

Image selection
Based on contact sheets with series negatives, volunteers choose from the best images of that series.

Tagging
Adding keywords to photos. This can be done with pre-conceived 'tags', but it is also possible to let volunteers create keywords.

1.3 Best results
If other tools are needed to unlock sources, there the programme gives space to tailor-made development. Research has shown that the quality that the crowd delivers to *Vele Handen* is unprecedented. The passion with which the volunteers work at *Vele Handen* is great. In addition, the programme includes technical measures to monitor quality. Of course, the first step of all is to provide a good manual for each project that can be downloaded before the participant starts working on the project. In addition, *Vele Handen* has been set up in such a way that the same scan is always entered by two different volunteers and then checked by a third person. If the transcriber has any questions about the scan, he or she will inform a designated expert of the heritage institution of the project who knows the collection well.

1.4 Involvement
By crowdsourcing a project, a large group of volunteers is working to make a collection accessible. They can be volunteers who are already working for the museum, collection or project, but also unknown people will join. Because the participants of the projects post their questions and comments on the forum, the museum gets a good idea of their concerns. And the participants are very happy with the help they get by the comments of the professionals. By making a project accessible online, involvement with the organization arises and the heritage institute/ museum gets to know their target group.

1.5 Fast, cheap and efficient
Indexing is extremely time-consuming. Having the work done by a large group of volunteers is therefore very efficient. And because the volunteers also like to make sources accessible, and can do this at their own time, projects go fast. Institutions also benefit from low start-up costs because it is a shared platform with other heritage institutions. They can decide themselves how much time they are working on the project themselves. This allows them to keep control. It is also possible to have the project supervised by Picturae from start to finish. In addition, paid entry is possible if the heritage institution does not want to use volunteers.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

1.6 Some figures*

Since the start of the project, 74 larger and smaller projects were set up at *Vele Handen*. Of those projects, 30 are still running, 44 projects have been accomplished. In total 4.996.994 scans have been uploaded. The biggest number of scans was for a project called *Alle Amsterdamse Akten* where people are asked to join on a voyage of discovery and help to rewrite the history of Amsterdam! In this project that started in September 2016, 704 people take part. The smallest project in terms of number of scans is *Kijk je Rijk!* (free translation: looking makes rich!) where people participate in the linking of data to 1700 video clips from the collection of the local broadcast Amersfoort. At first 190 scans were uploaded, 104 people participate. Later on, when it turned out to be successful, another 730 scans were uploaded in the next project *Kijk je Rijker!* (looking makes richer!) where 271 people take part.
For all projects 26.985 people joined *Vele Handen*, some of them just work on a small number of scans, but there are also people who have up to 20.000 scans or more on their account. In this figure we need to take into account that some people work on more projects at the same time. If you count unique transcribers, the number is around 15.000 participants

All scans are double entered by two different people, after that they are checked by, or the employees of the heritage institutions, or by specialists. Of the 4.996.994 scans, over the years 3.372.703 (76%) have been double entered and from those 3.208.454 (72%) have been checked.

Looking at all days of all projects during the runtime, choosing 24 January 2019 as closing day for the projects that are still running, it adds up to 52.898 days. This means that with 3.372.703 double entered files, the average of a day is 66 scans, for the checked files this is 62 per day.

Not all scans can be entered and or checked. From the completed projects, the average is 89% double entered and 88% of what also has been checked. Sometimes it does not work; the project *Post van Weldadigheid* was stopped when only 42% was finished.

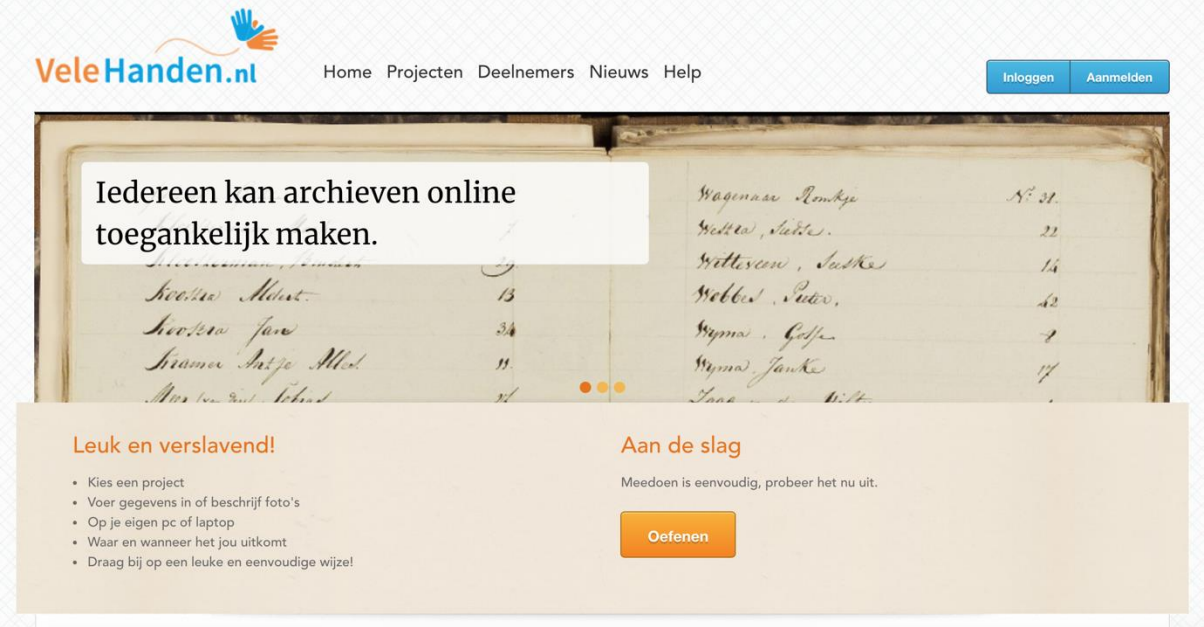
(*results of one project and runtime of one other project were not available, nevertheless this overview gives insight in the total).

2. How is Vele Handen presented to the public?

'Everyone can make archives accessible online' Is the key sentence used on the starting page for transcribers. 'It is fun and addictive'. Furthermore the transcriber reads the 5 steps it takes to participate.

- Choose a project
- Enter data or describe photos
- On your own PC or laptop
- Where and when it suits you
- Contribute in a fun and easy way!

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



Possible transcribers can practise before they make their own account. After that they are invited to join a project; 'It is just like a sudoku or a crossword puzzle. *Vele Handen*; addictive and fun in addition'.
There is a lot of attention for the title of a project, if you can find a way to make the collection appealing to a larger group of people, it is more likely that they will choose working on that collection. Nevertheless, there are certain projects that are very much linked to those who are directly involved, e.g. Missing Links is linking data and images of genealogical sources from Leiden and surroundings. Genealogy is booming, we all want to know where we come from, and who our ancestors were. This project generated over the runtime of 6 years 2411 participants. Now almost 85% of the scans have been worked. This project will come back in chapter 3, where two projects are more elaborated.

2.2 Experience is not necessary, only passion for the past

To become a volunteer, you don't need to be experienced on forehand. If interested in history, or with a passion for the past, people can become volunteer. Since the volunteers work from their own computer, they can work at a time that suits them best.

2.3 What is expected of the user?

The requested work can vary per project. In many cases it concerns the retyping of information in fields that can be seen next to the scan, for example: name, date, place of residence. But there are also projects where volunteers are asked to provide photos with a description. They can choose which project or projects they participate in. Making a mistake is not bad, since there is a second opinion. And there is a forum with questions, where other volunteers and archivists from the institutions help.

2.4 What does the user get in return?

There are participants who have 20,000 scans on their name. But there are also volunteers who like to occasionally enter a scan. It's all possible. And every effort is rewarded with *Vele Handen* points. Volunteers can hand in these points for a small fee when setting up the project, for example in the form of downloading scans or admission to an exhibition.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

3. Two projects

Volunteers from Militia-registers, the first project on *Vele Handen*, have indexed more than 300,000 scans within 18 months! Some top importers already have 20,000 scans on their name

3.1 How things started

Setting up The Militia-register project, the first project Picturae ran on *Vele Handen*.
The scans of the militia-register books have been placed on velehanden.nl. With each scan a form is shown on which the participants can enter the handwritten names. A pointer page has been created on *Vele Handen* where participants can see what militia registers are and how the entry is done.

3.2 Entering by the volunteers

The data from each scan is entered twice and then checked. This ensures the quality of the search system. Each volunteer earns points for entering data. In the case of Militia registers, he can use it to download scans for which he would otherwise have had to pay.
Heritage collections can decide for their self whether you want to give the volunteers a reward and for which they can redeem the accumulated points. The volunteers can exchange knowledge and experience on the forum of the website.

3.3 Exporting the data

As the final step, the verified data is exported. The Militia registers can be searched on www.militieregisters.nl. Heritage collections can determine self what you do with the data. Integrate within their own website, or present on a completely new site.

References

1. Newsroom. Green paper on Citizen Science for Europe: Towards a society of empowered citizens and enhanced research. *Digital Single Market - European Commission* (2014). Available at: https://ec.europa.eu/digital-single-market/en/news/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research. (Accessed: 9th May 2019)
2. Schade, Sven & Tsinaraki, Chrysi. Survey report: data management in Citizen Science projects. *EU Science Hub - European Commission* (2016). Available at: https://ec.europa.eu/jrc/en/publication/survey-report-data-management-citizen-science-projects. (Accessed: 9th May 2019)

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

3. Kullenberg, C. & Kasperowski, D. What Is Citizen Science? – A Scientometric Meta-Analysis. *PLOS ONE* **11**, e0147152 (2016).
4. Kosmala, M., Wiggins, A., Swanson, A. & Simmons, B. Assessing data quality in citizen science. *Front. Ecol. Environ.* **14**, 551–560 (2016).
5. Nelson Gil & Ellis Shari. The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20170391 (2019).
6. Ballard, H. L. *et al.* Contributions to conservation outcomes by natural history museum-led citizen science: Examining evidence and next steps. *Biological Conservation* **208**, 87–97 (2017).
7. Sarah Phillips, Mathias Dillen, Quentin Groom, Laura Green, Marie-Hélène Weech, Noortje Wijkamp. Report on New Methods for Data Quality Assurance, Verification and Enrichment.
8. Zooniverse. *Project Plumage, Zooniverse.* Available at: https://www.zooniverse.org/projects/ghthomas/project-plumage/about/research. (Accessed: 29th May 2019)
9. Cappa, F., Laut, J., Porfiri, M. & Giustiniano, L. Bring them aboard: Rewarding participation in technology-mediated citizen science projects. *Computers in Human Behavior* **89**, 246–257 (2018).
10. ECSA documents. *European Citizen Science Association (ECSA)* Available at: https://ecsa.citizen-science.net/documents. (Accessed: 24th July 2018)
11. 10 Principles of Citizen Science – Australian Citizen Science Association.
12. Alender, B. Understanding volunteer motivations to participate in citizen science projects: a deeper look at water quality monitoring. *JCOM* **15**, A04 (2016).

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

13. Conrad, C. C. & Hilchey, K. G. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environ Monit Assess* **176**, 273–291 (2011).
14. Hunter, J. & Hsu, C.-H. Formal Acknowledgement of Citizen Scientists' Contributions via Dynamic Data Citations. in *Digital Libraries: Providing Quality Information* (eds. Allen, R. B., Hunter, J. & Zeng, M. L.) 64–75 (Springer International Publishing, 2015).
15. Bird, T. J. *et al.* Statistical solutions for error and bias in global citizen science datasets. *Biol. Conserv.* **173**, 144–154 (2014).
16. Hochachka, W. M. *et al.* Data-intensive science applied to broad-scale citizen science. *Trends Ecol. Evol.* **27**, 130–137 (2012).
17. Jacobs, C. & Zipf, A. Completeness of citizen science biodiversity data from a volunteered geographic information perspective. *Geo-Spatial Information Science* 1–11 (2017). doi:10.1080/10095020.2017.1288424
18. Lukyanenko, R., Parsons, J. & Wiersma, Y. F. Emerging problems of data quality in citizen science: Editorial. *Conservation Biology* **30**, 447–449 (2016).
19. Snall, T., Kindvall, O., Nilsson, J. & Part, T. Evaluating citizen-based presence data for bird monitoring. *Biol. Conserv.* **144**, 804–810 (2011).
20. Belbin, L., Daly, J., Hirsch, T., Hobern, D. & LaSalle, J. A specialist's audit of aggregated occurrence records: An 'aggregator's' perspective. *ZooKeys* **305**, 67–76 (2013).
21. Veiga, A. K. *et al.* A conceptual framework for quality assessment and management of biodiversity data. *PLOS ONE* **12**, e0178731 (2017).
22. Allahbakhsh, M. *et al.* Quality Control in Crowdsourcing Systems: Issues and Directions. *IEEE Internet Computing* **17**, 76–81 (2013).

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

23. Projekte | Buerger schaffen Wissen. Available at: https://www.buergerschaffenwissen.de/projekte. (Accessed: 7th May 2019)
24. Welcome! | National Plant Monitoring Scheme. Available at: https://www.npms.org.uk/. (Accessed: 7th May 2019)
25. GBIF - Darwin Core Archive Assistant v1.1. Available at: http://tools.gbif.org/dwca-assistant/. (Accessed: 29th March 2019)
26. Chapman, A. D. *Principles of Data Quality*. (Global Biodiversity Information Facility, 2005).
27. Citizen Science Association Data & Metadata Working Group: Report from CSA 2017 and Future Outlook. *Wilson Center* (2017). Available at: https://www.wilsoncenter.org/article/citizen-science-association-data-metadata-working-group-report-csa-2017-and-future-outlook. (Accessed: 30th July 2018)
28. WG 5 Deliverable: Citizen Science Ontology | COST Action CA15212. Available at: https://www.cs-eu.net/news/wg-5-deliverable-citizen-science-ontology. (Accessed: 10th April 2019)
29. Project Finder | Australian Citizen Science Project Finder. Available at: https://biocollect.ala.org.au/acsa/#isCitizenScience%3Dtrue%26isWorldWide%3Dtrue%26max%3D20%26sort%3DnameSort. (Accessed: 23rd April 2019)
30. Le Bras, G., Chagnoux, S. & Dillen, M. *Specification of data exchange format for transcription platforms*. (Zenodo, 2019). doi:10.5281/zenodo.2598413
31. Citizen science. *Wikipedia* (2019).
32. About the Citizen Science Association | Association. *Citizen Science* Available at: http://www.citizenscience.org/association/about/. (Accessed: 25th July 2018)
33. Australian Citizen Science Association – Citizen science is redefining how we do science. Available at: https://citizenscience.org.au/. (Accessed: 25th July 2018)

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

34. European Citizen Science Association (ECSA). *European Citizen Science Association (ECSA)* Available at: https://ecsa.citizen-science.net/. (Accessed: 25th July 2018)
35. Australian Museum Centre for Citizen Science - Australian Museum. Available at: https://australianmuseum.net.au/australian-museum-centre-for-citizen-science. (Accessed: 25th July 2018)
36. Citizen Science Association. *Schoodic Institute* Available at: https://www.schoodicinstitute.org/what-we-offer/educational-scientific-partnerships/citizen-science-association/. (Accessed: 25th July 2018)
37. Concept Note: Citizen Science Global Partnership. *Wilson Center* (2018). Available at: https://www.wilsoncenter.org/article/concept-note-citizen-science-global-partnership. (Accessed: 30th July 2018)
38. COST Action CA15212. Available at: https://www.cs-eu.net/. (Accessed: 30th July 2018)

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483