Developing metrics and instruments to evaluate citizen science impacts on the environment and society

EC Horizon-2020 Grant Agreement number 824711

Call: H2020-SwafS-2018-2020 (Science with and for Society)

Topic: SwafS-15-2018-2019

Type of action: RIA

Deliverable 2.4: MICS corpus of knowledge

Delivery Year: 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824711.

Document Information

Project Number	824711			**Acronym**	MICS
Full title	Developing metrics and instruments to evaluate citizen science impacts on the environment and society				
Project URL	**www.mics.tools**				
EU Project officer	Yiannis Vacondios				
Deliverable	**Number**	D2.4	**Title**	MICS corpus of knowledge	
Work package	**Number**	2	**Title**	Methods for measuring citizen-science impact	
Date of delivery	**Contractual**	Month 14		**Actual**	March *2020*
Dissemination Level	Public				
Authors (Partner)	IHE Delft				
Responsible Author	Uta Wehn		**Email**	u.wehn@un-ihe.org	
	Partner	IHE Delft			
Abstract (for dissemination)	The MICS project develops approaches and tools to evaluate citizen-science impacts. These approaches and tools can help to plan and implement projects in ways that lead to more robust results. This report describes the structure and process of building the MICS corpus of knowledge (or knowledge base), which forms part of the MICS methodological and technical efforts to capture, analyse and consolidate insights on the impacts of citizen science.				
Keywords	Knowledge base, database, human-readable, machine-readable, metadata				

Version Log

Version as date	Author	Partner	Change
2020_03_05	Mohammad Gharesifard	IHE Delft	Initial document structure and draft content for section 2
2020_03_17	Uta Wehn	IHE Delft	Revision of structure. Update and addition of content throughout.
2020_03_20	Luigi Ceccaroni	Earthwatch	Edits throughout.
2020_03_23	Hannah Joyce	RRC	Review; minor edits / comments throughout.
2020_03_25	Uta Wehn	IHE Delft	Minor revisions based on reviewers' comments.
2020_03_27	Mohammad Gharesifard	IHE Delft	Minor revisions based on reviewers' comments and addition of Annex 2.
2020_03_31	Uta Wehn	IHE Delft	Final edits throughout; update of Annex 2

To cite this document:
Wehn, U., Gharesifard, M. and Ceccaroni, L. (2020). D2.4 MICS corpus of knowledge. *Deliverable report of project H2020 MICS (grant agreement No 824711)*

Table of Contents

Executive Summary

The MICS project develops approaches and tools to evaluate citizen-science impacts. These approaches and tools can help to plan and implement projects in ways that lead to more robust results.

This report is a deliverable of Work Package 2 (WP2) – 'Methods for measuring citizen-science impact'. The purpose of this report is to describe the structure and process of building the MICS corpus of knowledge (or knowledge base), which forms part of the MICS methodological and technical efforts to capture, analyse and consolidate insights on the impacts of citizen science. The report introduces the general concept of a corpus of knowledge before detailing the specific characteristics of the MICS corpus of knowledge.

A corpus of knowledge is a collection and consolidation of facts, information and insights acquired through experience or education and the theoretical or practical understanding of a specific topic. As such, a corpus of knowledge can be considered a knowledge base, which has more expressive relations than a database. The MICS corpus of knowledge serves to capture and describe knowledge in a specific field, namely knowledge on the impacts of citizen science in five domains: society, governance, the economy, the environment, and science. Specifically, this corpus of knowledge consists of a set of data, metrics and rules (in part defined and detailed in D2.2 and D2.3) as well as documents and indicators (to be reported in D2.7) about measuring citizen-science impact for strategies, policies, actions and innovative practices (especially in science as initial case-study, and later in all the five domains).

Setting up a corpus of knowledge follows generic steps (outlined in this document) that have been tailored for the process of building the MICS corpus of knowledge. Capturing and sharing project metadata is a fundamental part of the content of MICS corpus of knowledge. The project metadata used by MICS is based on previous efforts for producing such metadata such as European projects: COBWEB and WeObserve, and activities of Working Group 5 of the COST Action on Citizen Science for improving data standardisation and interoperability across citizen-science projects.

The MICS corpus of knowledge is created, developed and shared as scientific evidence for policy makers and practitioners, and as an example for other projects that need to build their knowledge base. It is part of MICS's inclusive, bottom-up approach in the sense that it does not only contain the work of a "handful of experts", but also involves the broader scholarly community and other stakeholders both at the beginning and throughout the whole process of handling the relevant research findings based on the principles of open science. The resulting consolidated corpus of knowledge will, as a result of this, not only provide added value regarding the plurality of its contained perspectives, but also benefits legitimacy-wise.

Additionally, the corpus of knowledge is inclusively created based on a continuous evaluation of needs implemented together with stakeholders, who, as a result of the evaluations, can further build up the corpus of knowledge under the guidance of WP2. Ideally, that ensures that the corpus of knowledge accurately reflects the views and needs of the targeted

audience. Thus, the comparison with and lessons learned from other projects and other methodologies will lead to a better understanding and recommendations for a more effective impact assessment of citizen science.

The corpus of knowledge will grow with the project, especially when the MICS platform becomes live and populated with projects' data. In summary, MICS will undertake the task to convert the data gathered in Phase 2 (pilot testing) into well-founded impact-assessment recommendations able to improve citizen science, and, in Phase 3 (recommendation) of the project, MICS will have a consolidated corpus of knowledge on the effectiveness and impact of citizen science, and this will also include knowledge in the form of recommendations and guidelines.

1 Introduction

1.1 Background on MICS

The MICS project develops approaches and tools to evaluate citizen-science impacts. These approaches and tools can help to plan and implement projects in ways that lead to more robust results.

The MICS project specifically aims to:

- provide comprehensive, participatory and inclusive metrics and instruments to evaluate citizen science impacts;
- implement an impact-assessment knowledge-base through toolboxes for methods application, information visualisation, and delivery to decision makers, citizens and researchers;
- improve the effectiveness of nature-based solutions through test-site development and citizen-science tool validation;
- generate new approaches that strengthen the role of citizen science in supporting research and development;
- foster a citizen-science approach to increase the extent to which scientific evidence is taken up by policy makers through recommendations and guidelines.

The result is an integrated platform where these metrics and instruments are available for use by anyone involved in a citizen-science project wanting to understand its impact, whether at the planning stage or several years after the project's conclusion. This platform is validated by pilot testing in four test and validation sites across Europe. These sites explore the applicability of MICS impact-assessment tools in regions with differing needs, contexts, and approaches to nature-based solutions, and with various levels of citizen-science application. For example, in Western Europe, river restoration is increasingly carried out within an ecosystem-based management framework at river or catchment scale; in Southern Europe, river restoration tends to be issue-specific with some ecosystem relevance; in Central and Eastern Europe, river restoration is about ecosystem protection and related to existing infrastructure. The four test and validation sites selected are in the UK, Italy, Hungary and Romania.

1.2 Purpose and scope of this report

This report is a deliverable of Work Package 2 (WP2) – 'Methods for measuring citizen-science impact'. The purpose of this report is to describe the structure and process of building the MICS corpus of knowledge, which forms part of the MICS methodological and technical efforts to capture, analyse and consolidate insights on the impacts of citizen science (see Figure 1).

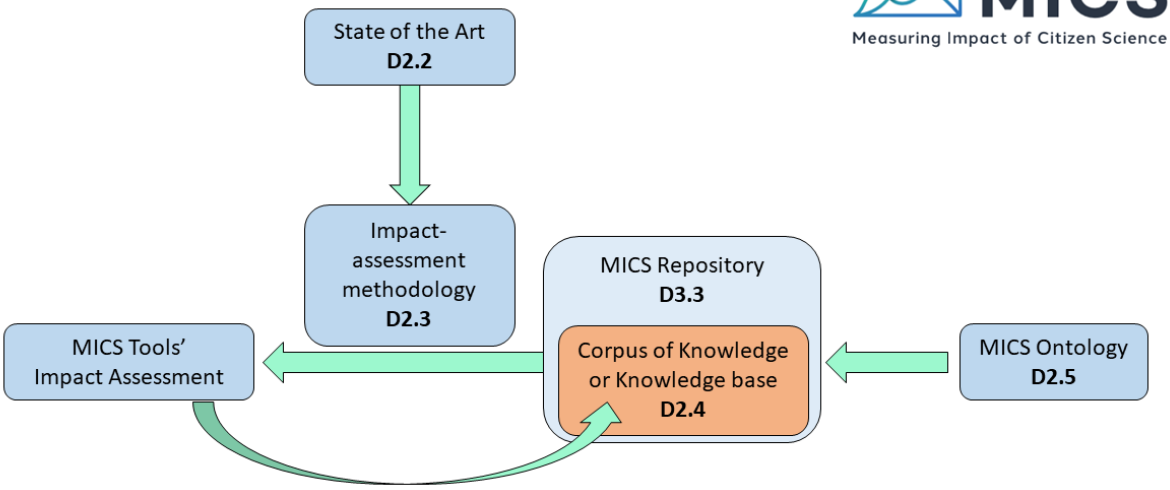

Figure 1. Structure of WP2 - 'Methods for measuring citizen science impact' and links with Work Package 3.

This document focuses on deliverable 2.4 'corpus of knowledge (knowledge base)' a key component informing the development of the MICS tools.

Specifically, the MICS **corpus of knowledge** or **knowledge base** is hosted within the **MICS repository (detailed in deliverable D3.3)**. Both the corpus of knowledge and the MICS repository's concepts, relations and structure are defined in the **MICS ontology (described in deliverable D2.5)**. The corpus of knowledge uses content from the **MICS impact assessment methodology (described in deliverable D2.3)** to inform the algorithms and tools of the MICS platform. The data collected by the MICS tools is fed back to the MICS repository to augment the knowledge base.

1.3 Structure of the report

This report is structured as follows. Following this introductory chapter, section 2 introduces the general concept of a corpus of knowledge before detailing the specific characteristics of the MICS corpus of knowledge. Section 3 concludes with an indication of which MICS tasks will capitalise on the results of this report as well as the next steps for MICS.

2 MICS corpus of knowledge

2.1 Introduction

This section describes the MICS corpus of knowledge, both in terms of its structure as well as the process of producing the content of the corpus of knowledge. We start the section with a general introduction and definition of what constitutes a corpus of knowledge.

2.1.1 What is a corpus of knowledge?

From a knowledge management perspective, a corpus of knowledge is a collection and consolidation of facts, information and insights acquired through experience or education and the theoretical or practical understanding of a specific topic. As such, a corpus of knowledge can be considered a knowledge base which is more comprehensive and more expressive than a database, as illustrated by the respective definitions:

- A database is a collection of data that is stored in a computer and that can easily be used and added to (Cox et al., 1997).
- A knowledge base is not a static collection of information; rather it is "a form of computerised database which is used as a store for knowledge on a given topic or range of topics" (Miller & Hannam, 1985).

Importantly, in this context, a knowledge base or corpus of knowledge consists of *codifiable* knowledge, i.e. knowledge that can be articulated, expressed, shared and stored digitally as compared to *tacit* knowledge (Polanyi, 1958; 1966) which is challenging to capture, articulate and share other than through common experiences or joint knowledge co-creation mechanisms.

A corpus of knowledge can be classified as human-readable or machine-readable (Molina & Blasco (2008)

- In a human-readable knowledge base, the content "must be comprehensible by readers who are not specialised in computer languages. Users should be able to manually perform the described tasks by reading the instructions of the document" (Molina & Blasco, 2008, p.3).
- In a machine-readable knowledge base, the content "must be able to be interpreted by a program that can perform automatically the tasks described in the document" (Molina & Blasco, 2008, p.3).):

2.1.2 Description of the MICS corpus of knowledge

The MICS corpus of knowledge serves to capture and describe knowledge in a specific field, namely **knowledge on the impacts of citizen science in five domains: society, governance, the economy, the environment, and science**.

2.2 Process of building the MICS corpus of knowledge

Setting up a corpus of knowledge follows generic steps (outlined below) that have been tailored for the process of building the MICS corpus of knowledge. Capturing and sharing project metadata is a fundamental part of the content of MICS corpus of knowledge. The project metadata used by MICS is based on previous efforts for producing such metadata such as European projects: COBWEB and WeObserve, and activities of Working Group 5 of the COST Action on Citizen Science for improving data standardisation and interoperability across citizen-science projects.

The steps for setting up the MICS corpus of knowledge are as follows.

1. **Agree on the structure and content types of the knowledge base**

   The structure for the MICS knowledge base is based on the MICS ontology (detailed in MICS deliverable D2.5).

   a. *Project metadata*
      Metadata presented in Annex 1.
   b. *Project data*
   c. *Project documents*
   d. *Rules for capturing the impact*
      Based on the impact domains identified in by D2.2 and the indicators to be detailed in D2.3

2. **Add content to the knowledge base**

   Content will be added via a variety of mechanisms and sources, including:

   - manually by MICS personnel;
   - via API connecting to existing project repositories;
   - as inputs from the MICS case studies;
   - by project coordinators, academics and citizen science practitioners from other citizen-science projects using the MICS platform.

3. **Publish the knowledge base**
   - Access will be provided on the MICS platform with careful consideration of how to make this available and easily accessible (e.g. via online reports, selected content with/without login, etc.).

4. **Analyse and improve the knowledge base**

   Analysis and consistency check of the content will be done continuously during the MICS project.

Reflection, updates and additions in terms of structure and content stem from the application of the MICS impact assessment methodology in the MICS case studies as well as other citizen science projects via the MICS platform.

Capturing and sharing project metadata is a fundamental part of MICS. The project metadata presented in Annex 1 is based on previous efforts for producing such metadata. More explicitly, the metadata shown in Annex 1 is built on efforts in other projects and activities for improving data standardisation and interoperability across citizen science projects.

In 2018, the WeObserve project published a landscape report of Citizen Science projects (Gold, 2018). This report includes a Project Description Template from the COBWEB project that is designed for capturing multiple choice and descriptive information about Citizen Science projects.

The COST Action CA15212 "Citizen Science to promote creativity, scientific literacy, and innovation throughout Europe" (2016-2020)' working group 5 to "Improve data standardisation and interoperability" has also been particularly active in producing a Core Conceptual Model (CCM) for citizen-science projects. This working group produced and presented an ontology of citizen science in the form of changes and extensions to the PPSR conceptual model (Version 1) in the Citizen Science Association Data & Metadata Working Group: Report from CSA 2017 and Future Outlook (Bowser et al., 2017). In 2018, the COST ACTION working group 5 produced a Citizen Science Ontology report (Ceccaroni et al., 2018), that aimed at initiating discussions for a next version of the aforementioned PPSR CCM.

The most recent instance of developing a metadata template that is applicable across Citizen Science projects is the example from the EU-Citizen.Science. This template that is presented in Annex 1 builds on the efforts above for developing such categorisations, and will be adopted by MICS as a basis for input project information in the MICS corpus of knowledge. It is important to note that this list of project metadata has not (yet) been published by the EU-Citizen.Science and therefore the reference given here is to the EU-Citizen.Science Zenodo page[1].

An example of input content for the MICS corpus of knowledge is the insights generated during a workshop in May 2019 at the River Restoration Conference in Liverpool. This workshop was held with practitioners and researchers from the field of NBSs and other areas to gain insights about impact evaluation approaches and indicators. These insights are reported by Novoa Gautier (2019) in a MICS-internal document. Another example of input for the MICS corpus of knowledge consists of the entries generated from in-depth interviews that were conducted in February and March 2020 with potential end-users of the MICS tools to understand their requirements (see Annex 2).

The MICS corpus of knowledge is created, developed and shared as scientific evidence for policy makers and practitioners, and as an example for other projects that need to build their

[1] See: https://zenodo.org/communities/eu-citizenscience/?page=1&size=20

knowledge base. It is part of MICS's inclusive, bottom-up approach in the sense that it does not only contain the work of a "handful of experts", but also involves the broader scholarly community and other stakeholders both at the beginning and throughout the whole process of handling the relevant research findings based on the principles of open science. The resulting consolidated corpus of knowledge will, as a result of this, not only provide added value regarding the plurality of its contained perspectives but also benefits legitimacy-wise.

Additionally, the corpus of knowledge is inclusively created based on a continuous evaluation of needs implemented together with stakeholders, who, as a result of the evaluations, can further build up the corpus of knowledge under the guidance of WP2. Ideally, that ensures that the corpus of knowledge accurately reflects the views and needs of the targeted audience. Thus, the comparison with and lessons learned from other projects and other methodologies will lead to a better understanding and recommendations for a more effective impact assessment of citizen science. The corpus of knowledge will grow with the project, especially when the MICS platform will be live and populated with projects' data.

3 Conclusions

This report has detailed the structure and process of building the MICS corpus of knowledge (or knowledge base), which forms part of the MICS methodological and technical efforts to capture, analyse and consolidate insights on the impacts of citizen science.

Next steps for MICS:

- In July 2020 (M19), the MICS consolidated corpus of knowledge will be formally integrated within the MICS repository when the project will reach milestone MS2.
- When the MICS repository (including the corpus of knowledge) will be integrated into the MICS platform (WP3), it will then be accessible by end-users (see Figure 1).
- Later in the project, WP4 will try to demonstrate/validate the usability of the consolidated corpus of knowledge for the expert community dealing with research and innovation, policy, implementation and education.
- Based on the consolidated corpus of knowledge, the MICS partners will produce factsheets covering the most pressing issues for each of the three regions.

In Phase 3 of the project, MICS will have a consolidated corpus of knowledge on the effectiveness and impact of citizen science; and this will also include knowledge in the form of recommendations and guidelines: MICS will undertake the task to convert the data gathered in Phase 2 into well-founded, hands-on impact-assessment recommendations able to improve citizen science.

References

Bowser, A., Brenton, P., Stevenson, R., Newman, G., Schade, S., Bastin, L., . . . Oliver, J. (2017). Citizen Science Association Data & Metadata Working Group: Report from CSA and Future Outlook, Washington, DC.

Ceccaroni, L., Schade, S., Bastin, L., Tsinaraki, C., Lemmens, R., Falquet, G., . . . De La Riva, S. (2018). D5.1 Citizen-science ontology. Deliverable report of project H2020 COST Action CA15212 "Citizen Science to promote creativity, scientific literacy, and innovation throughout Europe". DOI: 10.5281/zenodo.3721074.

Cox, I. J., Miller, M. L., Omohundro, S. M., & Yianilos, P. N. (1997). Multimedia database retrieval system which maintains a posterior probability distribution that each item in the database is a target of a search: Google Patents.

Gold, M. (2018). WeObserve deliverable D2.1, EU Citizen Observatories Landscape Report - Frameworks for mapping existing CO initiatives and their relevant communities and interactions.

Miller, A. S., & Hannam, R. G. (1985). Computer Aided Design Using a Knowledge Base Approach and its Application to the Design of Jigs and Fixtures. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 199(4), 227-234. DOI:10.1243/PIME_PROC_1985_199_073_02

Molina, M., & Blasco, G. (2008). Human-Readable and Machine-Readable Knowledge Bases Using Specialized Word Processors. Paper presented at the 2008 20th IEEE International Conference on Tools with Artificial Intelligence.

Novoa Gautier, S. (2019) Milestone MS2.4. Workshop with practitioners and researchers from NBSs and other areas to validate the workflow for impact evaluation. Deliverable report of project H2020 MICS (grant agreement No 824711)

Polanyi, M. (1958). Personal Knowledge: Towards a Post-Critical Philosophy. University of Chicago Press, Chicago.

Polanyi, M. (1966) The Tacit Dimension. New York: Doubleday.

Annex 1 Metadata template based on EU-Citizen.Science PPSR CCM Metadata

PROJECT NAME

Origin UID	*Globally Unique Identifier in the database where a project was first registered. Allows traceability of a project in multiple databases to its original registration. For example, the Grant Number.*
Origin Database	*The name of the project database where a project was first registered. Allows traceability of a project in multiple databases to its original registration. For example, the CORDIS registration.*
Project Name	*Short name or title of the project*
Project Aim	*Primary aim, goal or objective*
Description	*Abstract or description of the project*
Keywords	Keywords (comma separated) which are indexed and aid in searching for and finding projects.
Status	*The activity status of the project. Vocabulary = Not yet started / Active / Periodically active / On hold / Completed / Abandoned/Terminated*
Start Date	*DD/MM/YYYY*
End Date	*DD/MM/YYYY*
Duration	*Alternative to the Start and End Date, when the duration is known but the start date is not yet set.*
Project Topic	*Vocabulary = Agriculture & Veterinary science / Animals / Archaeology & Culture / Astronomy & Space / Biodiversity / Biogeography / Biology / Birds / Chemical sciences / Climate & Weather / Ecology & Environment / Education / Food science / Genetics / Geography / Geology & Earth science / Health & Medicine / Indigenous culture / Information & Computing sciences / Insects & pollinators / Long-term species monitoring / Ocean, Water, Marine & Terrestrial / Nature & outdoors /*

	Natural resource management / Physics / Psychology / Science policy / Social sciences / Sound / Transportation / Other
Image	*An image to represent a project (url to jpeg or png file)*
Image Credit	*A credit for the image used to represent a project*
Project URL	*URL to an external web site for the project*
Project Latitude	*Latitude coordinate of the center of the project area. Typically this is where the project is hosted, e.g., a home institution. https://www.latlong.net/convert-address-to-lat-long.html*
Project Longitude	*Longitude coordinate of the center of the project area. Typically this is where the project is hosted, e.g., a home institution. https://www.latlong.net/convert-address-to-lat-long.html*
Project Host / Coordinator	*Name of the primary organisation responsible for hosting or implementing the project.*
Public Contact	*Person that interested public or researchers should contact*
Public Contact Email	*Public contact email address*
Funding Program	*Indication of the program that funds or funded a project.*
Funding Source	*Sponsor(s) of the project*
How to Participate	*Free text description of how people can get involved in the project. Textual instructions for joining the project*
Participation Tasks	*Vocabulary:*
Equipment	*Required or suggested equipment*

Annex 2 Sample entries in the MICS Corpus of Knowledge

MICS partners Earthwatch, IHE and RRC, prepared and conducted eleven in-depth interviews with potential end-users of the MICS tools (i.e. project coordinators of citizen science projects) during February and March 2020 in order to better understand their specific requirements. In total, ten projects were covered, see Table below.

No.	Representative from project	Interview date
1	I-mars.eu	2020.02.20
2	Citclops	2020.02.24
3	FreshWater Watch	2020.02.24
4	Naturehood	2020.02.27
5	D-NOSES	2020.03.05
6	Earth Challenge 2020	2020.03.16
7	Swedish Mass Experiment 2020	2020.03.20
8	Cities-Health	2020.03.23
9	Outfall Safari	2020.03.27
10	ACTION	2020.03.27

The insights generated from these interviews present an example of input information for the MICS corpus of knowledge.

Themes	Sub themes	Example interview		
		Jobs	Pains	Gains
Purpose	"WP"	Had a WP on project website effectiveness		
	Project proposal			
	Reporting			
	Project learning		If you can't measure impact your project will fail, can't iterate on your design	Impact assessment helps to optimise project design to maximise data produced
	Promotion/communications			
	Interest	Research and personal		
Measuring	Methods used	Did not use a formal method	Not aware of methods	
	Effectiveness of methods	Confidence in the methods not important as nobody will double check your figures.	In the long-term unable to capture the broader/indirect impacts	In the short-term, provides a good idea of impact
	Ease of methods	It wasn't hard, it just wasn't something we had to do. Only done out of personal/professional interest	A lot of environmental/physical scientists find impact assessment difficult/not intuitive	MICS tools should be accessible regardless of background; if there was a tool that standardised impact assessment it might help people consider impact in a way they haven't before.
	Stage of project measured	Considered from the project design (written into proposal). Only considered at the end when reporting to the funder.	Many people are unlikely to consider impact during the project design.	Should be considered in the design of the project
	Time spent	Project with impact as WP: 100 hours; project without impact WP: 1 hour.	Time available in the project limits what can be done with impact assessment.	

Themes	Sub themes	Example interview		
		Jobs	Pains	Gains
Domains	Environment			
	Economy			
	Governance			
	Science	Interested in measuring academic impact		
	Society	Not interested in measuring public impact		
Indicators	Data type	Quantitative predominantly; but some user	Only measured what was readily available	
	Data points	Measured		
	Citizens engaged	Measured; and volunteer retention		
	Events organised			
	Volunteer preferences	Measured; with qualitative questionnaires		
	Geographic scope	Measured number of countries where volunteers uploaded		Good for converting to promotional map
	Publications	Not measured		
	Education (for participants)		Didn't have time to measure (wasn't part of the work-package)	Good for demonstrating what was given back to participants.
	RRI			
	Data quality			
	Case study/impact story			
Audiences	General comms	Radio interviews and university press releases		
	Specific scientific field	Reporting on research outputs of project; 6th (interest in impact assessment)		Good impact assessment shows citizen-science utility
	Participants	1st (interest in impact assessment)		

Themes	Sub themes	Example interview		
		Jobs	Pains	Gains
	Citizen-science community	2nd (interest in impact assessment)		
	Project team	2nd (interest in impact assessment)		
	Lead organisation (of project)	4th (interest in impact assessment)		
	Industry			
	Project funder	Need to justify spending public money; 4th (interest in impact assessment)		
Output	Map	For promotional material		Engaging, attractive on promotional material
	Score			Easy to interpret
	Qualitative data			Validates and adds context to any score
	Raw data option	CSV file available to download		Increases transparency of calculations, people trust the platform more even if they don't test the data
	Comparison	Not helpful to compare with other projects, impact assessment is too unique and cannot be compared.		
	Explanation			
	Recommendations			
	Terminology	Did not use impact; used knowledge creation for scientific discipline, not participants		
	Use of report	Outputs for personal interest and funder/box-ticking are different.		

Themes	Sub themes	Example interview		
		Jobs	Pains	Gains
	Good examples of impact assessment	field of internet privacy/legal issues have good online tools where you can self-assess your impact		
	Accessibility	Need to balance the users who are already interested in impact versus those who just want a tick-box exercise.		
Platform features	Endorsement from other users	Especially prominent members of the citizen science community		
	Endorsement by funders			
	Usable	need a simple and clear interface		
	Methodological rigour	Not too important, will only be scrutinised by those who score badly in the assessment		
	Design			