Measuring Impact of Citizen Science

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 2.5: A Multilingual Ontology

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**	D2.5	**Title**	*A Multilingual Ontology*
Work package	**Number**	2	**Title**	Methods for Measuring Citizen Science Impact

Date of delivery	**Contractual**	Month 14 (*February 2020*)	**Actual**	Month 14 (*February 2020*)
Dissemination Level	Public			

Authors (Partner)	Earthwatch		
Responsible Author	James Sprinks	**Email**	jsprinks@earthwatch.org.uk
	Partner	Earthwatch	

Abstract (for dissemination)	The multilingual ontology builds on the metadata scheme introduced in deliverable 3.3 “The MICS Repository”, and is presented as an associated module describing a project’s impact. It is related to Sustainable Development Goal (SDG) domains, and the ontology’s vocabulary will be based, where possible, on EuroVoc, the EU’s multilingual thesaurus [http://eurovoc.europa.eu/drupal/]. This effort to make the ontology multilingual is to ensure the ontology is accessible across the MICS case-study sites, and further afield.
Keywords	Metadata scheme, Ontology, Multilingual

Version Log

Version as date	**Author**	**Partner**	**Change**
14/02/2020	James Sprinks, Luigi Ceccaroni	Earthwatch	Document creation and intermediary version
21/02/2020	James Sprinks	Earthwatch	First draft completed
21/02/2020	James Sprinks	Earthwatch	Section 2.3 added, relating the MICS ontology and its potential applicability to other online citizen science platforms
28/02/2020	Luigi Ceccaroni	Earthwatch	Final review
05/03/2020	Mohammad Gharesifard	IHE Delft	Internal review: Minor revisions regarding terminology
12/03/2020	James Sprinks	Earthwatch	Addressed internal reviewer comments

To cite this document:

Ceccaroni, L. and Sprinks, J. (2020). D2.5: A Multilingual Ontology. *Deliverable report of project H2020 MICS (grant agreement No 824711)*

Contents

1 Executive Summary

The MICS multilingual ontology is a metadata scheme describing the structure of citizen-science project metadata, how they interact, and the position of impact assessment data in that fabric. It has been designed as a multilingual scheme, informed by the EuroVoc dictionary, to ensure it is easily translatable across the prominent languages of the test and validation locations. It informs the MICS repository, ensuring that standardised and interoperable data are used in its architecture.

2 Introduction

2.1 The context within the MICS project

The purpose of the MICS project is to investigate how citizen science adds value to research and innovation, and in doing so better understand the opportunities to improve this process. The project will develop several methods and procedures to measure citizen science impact, modified to be fit for purpose and involving completely original impact assessment indicators. To this end, the MICS project will collect a range of different data, both qualitative and quantitative, from several sources.

Deliverable 3.3 describes the repository where this data is collected and stored, along with a metadata scheme describing the relationship of the data within a citizen science project. However, the MICS project will also output new types of data relating to impact and how it is assessed. In this deliverable, a MICS impact ontology is presented, showing how impact data fits into the overall data scheme of a citizen science project, and how it relates to and can affect / be affected by other data types. The ontology will also link the impact data created by the MICS project to broader agendas, specifically the UN's *sustainable development goals* (SDGs)[1]. This project aims to contribute to significant knowledge and implementation gaps to more effectively integrate *natural-based solutions* (NBSs), citizen science and open-science methods. To this end, the following SDGs have been identified as the foundation for the implications of citizen science on NBSs:

- **Goal 3: Good Health and Wellbeing**
  Ensure healthy lives and promote well-being for all at all ages
- **Goal 4: Quality Education**
  Ensure inclusivity and equitable quality education and promote lifelong learning opportunities for all
- **Goal 6: Clean Water and Sanitation**
  Ensure availability and sustainable management of water and sanitation for all
- **Goal 11: Sustainable Cities and Communities**
  Make cities and human settlements inclusive, safe, resilient and sustainable
- **Goal 12: Responsible Consumption and Production**
  Ensure sustainable consumption and production patterns
- **Goal 15: Life on Land**
  Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat deforestation, and halt and reverse land degradation
- **Goal 16: Peace, Justice and Strong Institutions**
  Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels

To ensure that the MICS framework is relevant to a wide range of citizen science initiatives, test and validation sites have been identified to explore common challenges in different contexts. These sites use different approaches across a variety of European settings. They include:

- Western Europe (the UK, ecosystem restoration and ecosystem-based management);

[1] The United Nations Sustainable Development Goals Knowledge Platform, https://sustainabledevelopment.un.org/?menu=1300, accessed 17th February 2020

- Southern Europe (Italy, ecosystem restoration and issue-specific ecosystem-related actions); and
- Central and Eastern Europe (Hungary, ecosystem protection and infrastructure-related interventions).
- Eastern Europe (Romania, ecosystem protection, ecotourism and flood-risk reduction).

Therefore, the MICS ontology has to be easy to translate across several languages, especially the most common languages of the aforementioned case-study sites. This translation will help to ensure the ontology and the MICS impact data associated with it are understandable and applicable to as wide a range of use-cases as required. The MICS ontology will use the EuroVoc[2] dictionary as a basis to ensure that the terms used are multilingual, easy to translate and have a well-defined meaning.

2.2 Context within WP2

The focus of WP2 is to define the criteria to measure the impacts of citizen science in each target region. These criteria include meeting the demands and needs of the socio-economic and scientific context of those countries. These criteria will be incorporated into an adaptive framework that will serve as a foundation for further inquiries in the target regions. The MICS conceptual framework is based on small-scale impact analysis approach, based on a specified set of criteria to evaluate the relevance of impacts of citizen science towards NBS science in the target regions. Those criteria will feed into the framework that will serve as a coordinate system for the large-scale measuring of citizen-science impact and be easily translatable into the languages of the pilot countries. The culmination of this process will be to transform these collected impact data into a consolidated 'corpus of knowledge'; a knowledge base hosted on the MICS repository described in deliverable D3.3. Figure 1 gives an overview of how the deliverables of WP2 contribute to the corpus of knowledge, and how that, in turn, this informs the impact assessment methodology and MICS impact tools.

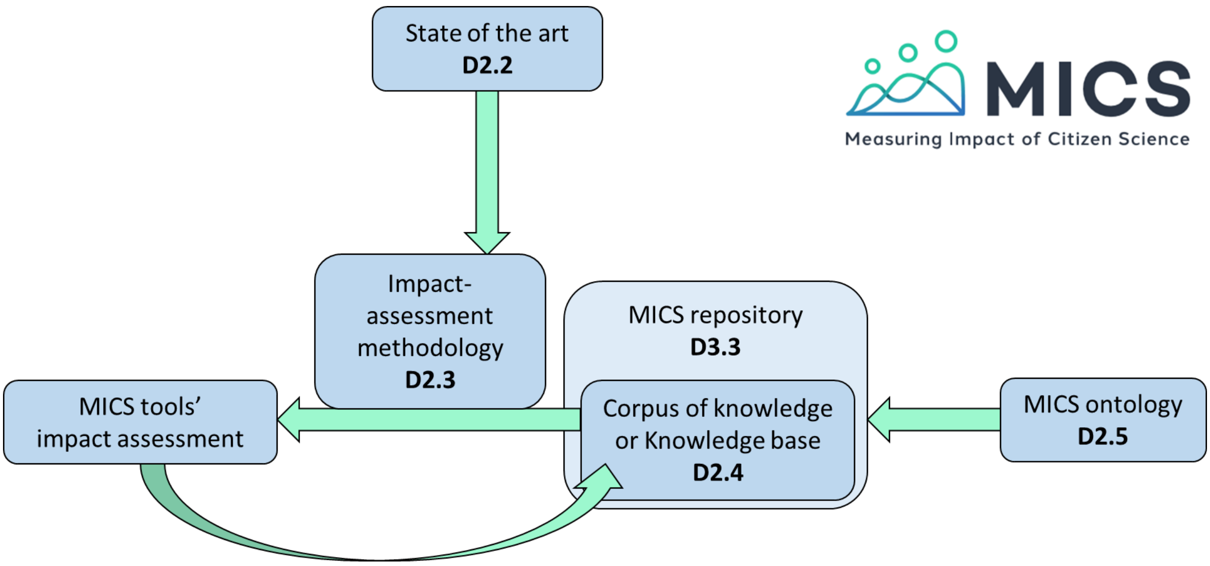


Figure 1. Overview of the MICS deliverables and relationships related to the ontology

[2] EU Vocabularies, https://op.europa.eu/en/web/eu-vocabularies, accessed 17th February 2020

As can be seen in figure 1, the MICS ontology (D2.5) forms a vital part of the repository (D3.3) describing the data structure of citizen-science projects, how the data are related, and the role that impact and its associated data play. The ontology also informs the corpus of knowledge (D2.4) and its structure, which then helps create the methodology for impact assessment (D2.3). From this methodology, a set of MICS tools are designed for impact assessment, allowing citizen-science practitioners to assess the impact of their project. The results of this process feed back into the MICS corpus of knowledge, forming an ever-evolving system, learning from the impact of existing and previous citizen-science practices and, in turn, informing future citizen science endeavours.

2.3 Broader Applications of the MICS Ontology

While the core aim of this deliverable and the ontology associated with it is concerned with the MICS project framework and in assisting in ensuring the outcomes regarding impact are meaningful, understandable and translatable across a range of citizen science contexts, it has the potential to be used across several other citizen-science collaborative initiatives.

For example, the EU Horizon 2020 Framework Programme (no. 776740) project WeObserve[3] "*aims to improve the coordination between existing Citizen Observatories and related regional, European and International activities. Its mission is to create a sustainable ecosystem of Citizen Observatories that can systematically address these identified challenges and help to move citizen science into the mainstream".* To improve the coordination between existing citizen observatories, one of the critical processes is to **map** existing citizen observatories, their relevant communities and relationships. A citizen science ontology could aid in this task, ensuring there is a consistent, intuitive and interoperable framework to follow.

Another example is the EU Horizon 2020 Framework Programme (no. 824580) project EU-Citizen.Science, which aims "*to build a central platform for citizen science in Europe. It will be a place to share useful resources about citizen science, including tools and guidelines, best practices and training modules. This will make citizen science knowledge created in Europe accessible to all and enable people to initiate their own activities. It will also enable anyone involved with or interested in citizen science to learn more and get involved".* Again, by utilising a metadata scheme such as the MICS ontology, a greater understanding of the data types, structure and their relationships can be achieved, ensuring the tools, guidelines and training developed are as widely applicable and usable as possible.

3 The MICS Ontology

3.1 Ontology background

As described in D3.3 "The MICS repository", the MICS ontology follows the Geneva Declaration on Citizen Science Data and Metadata Standards (Geneva, Switzerland, 6th June 2018[4]), and the ontology defined in the COST Action Working Group 5. In this deliverable, we build on that ontology, adding classes and properties associated with impact data, its assessment, and its relationship to the SDGs listed in the introduction.

[3] The WeObserve project, https://www.weobserve.eu/, accessed 21st February 2020

[4] Minutes of WG5 workshop in Geneva: "On the citizen-science ontology, standards & data", https://cs-eu.net/sites/default/files/media/2018/06/COST-WG5-GenevaDeclaration-Report-2018.pdf, accessed 18th February 2020. A description of how to access the OWL Code and ontology layout is available in the Appendix.

A more thorough description of the terminology, rationale and concepts of the core ontology can be found in D3.3. However, it is worth recapping for clarity that in the following description a “class” refers to what can be described as an “entity”, and “property” refers to what can be described as a “relationship”. In general, the MICS ontology follows the OWL 2[5] standard in its structure. It is also worth reiterating that the MICS metadata scheme and resulting ontology are inspired by but by no means formally connected to other ontologies, not committing to their semantics or inheriting any other specific conventions. For the following graphical descriptions of the MICS ontology, figure 2 illustrates the colour coding that will be used.

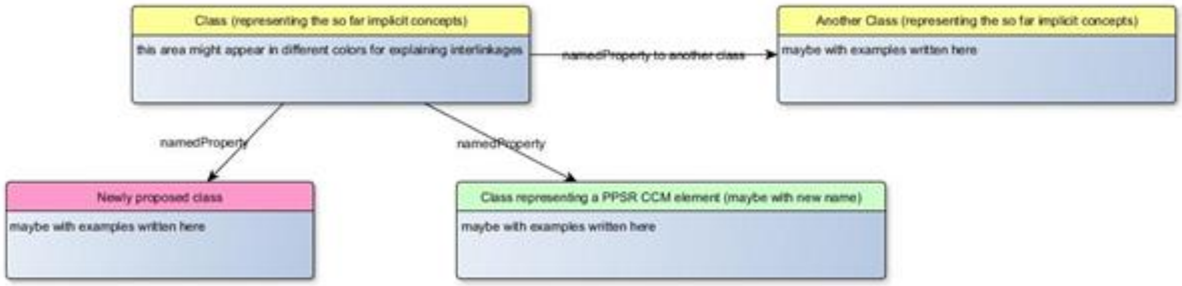
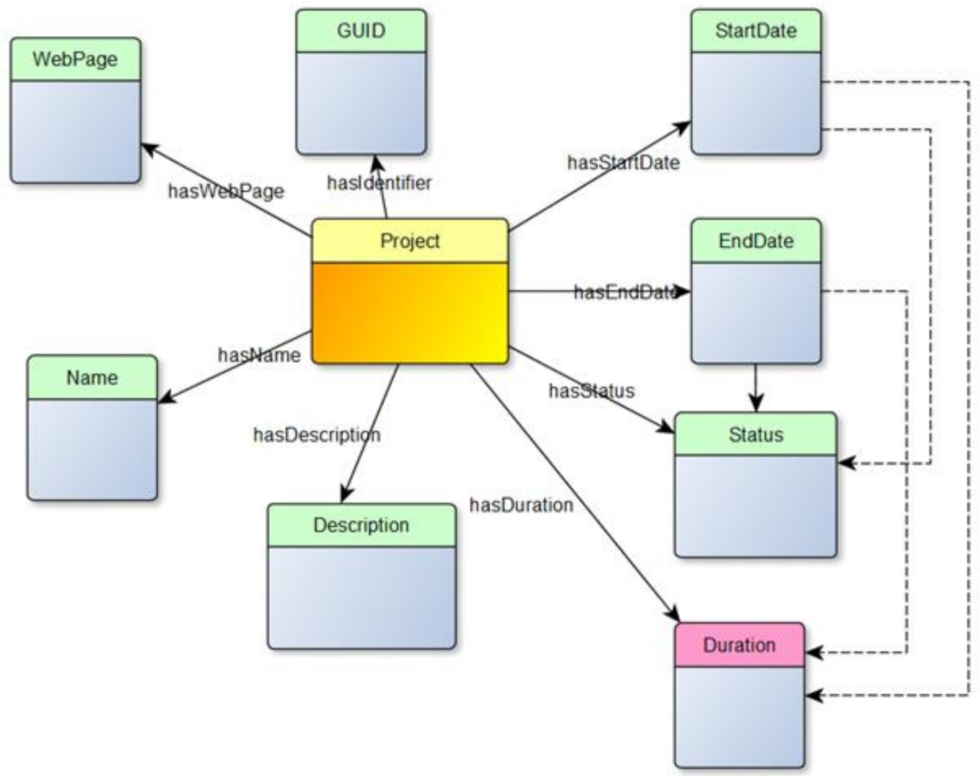


Figure 2. Colour coding of MICS ontology structure diagrams. *Classes representing concepts from the COST ontology are in yellow and green; newly proposed categories, added to the original ontology and related explicitly to the MICS project, are in pink.*

3.2 MICS Ontology Core

The MICS ontology is designed to be used in its entirety as a single structure; however, for clarity, its description will be presented in the form of representative modules. Each module will contain entities that are grouped because of their close, intuitive properties and relationships. (For instance, the Project Image module described in D3.3 describes if a project has an image, what credit is associated with it, if the image is a logo, etc.) Figure 3 below represents the Project Core of the MICS ontology: its classes and their associated properties. The dotted lines towards the right of the diagram, relating to temporal entities of a project, describe the relationships between them. For instance, the duration can be defined by the start and end date, and the status (is the project live, finished, etc.?) can also be inferred.

[5] OWL 2 Web Ontology Language Document Overview (Second Edition), https://www.w3.org/TR/owl2-overview/, accessed 18th February 2020

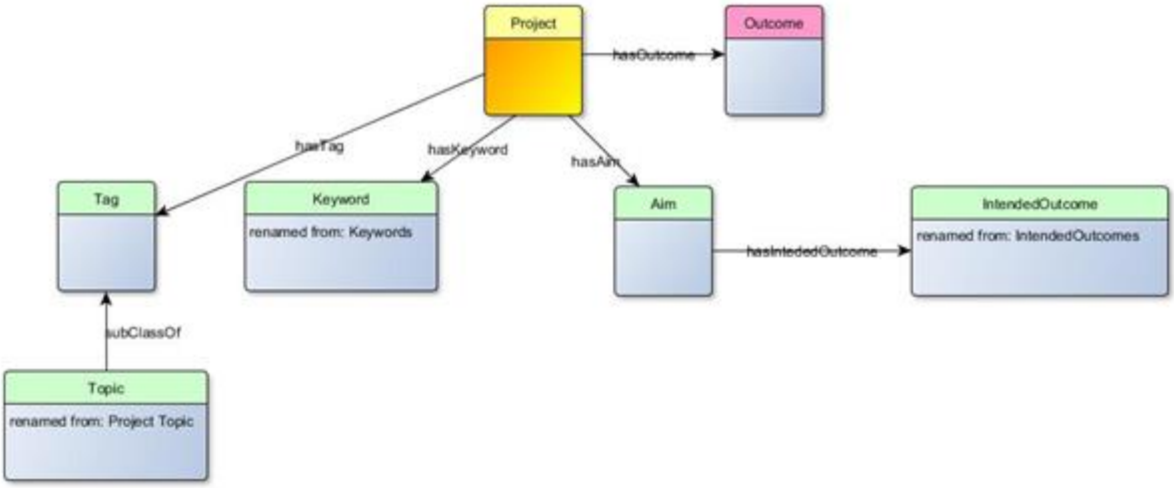


Property	**Description/comments**
hasDescription	It associates a project to its textual description.
hasDuration	It associates a project to its duration.
hasEndDate	It associates a project to its end date.
hasIdentifier	It associates a project to its *globally unique identifier* (GUID).
hasName	It associates a project to its name.
hasStartDate	It associates a project to its start date.
hasStatus	It associates a project to its activity status.
hasWebPage	It associates a project to its web page or URL.

Figure 3. The MICS ontology core classes and properties

3.3 Project Detailed Information

In deliverable D3.3, the MICS ontology's Project Detailed Information module was introduced. In terms of the MICS project aims regarding measuring and assessing impact, -----the classes "Outcome" and "IntendedOutcome" are key concepts. It is perhaps intuitive, at least in a first instance, to relate a projects' impact with its outcomes. For example, a project might have an outcome of reducing water wastage, which then might have a knock-on policy impact of the associated water board changing their policy on leak repair. This policy change could then profoundly impact the local community economically in terms of costs and environmentally in terms of the ecology of the area. However, it is worth reiterating that the MICS ontology and the corpus of knowledge it contributes to is an evolving framework, so there is scope for this to change as new experience is gathered (as shown in figure 2). Figure 4 describes the Project Detailed Information ontology module and associated properties.



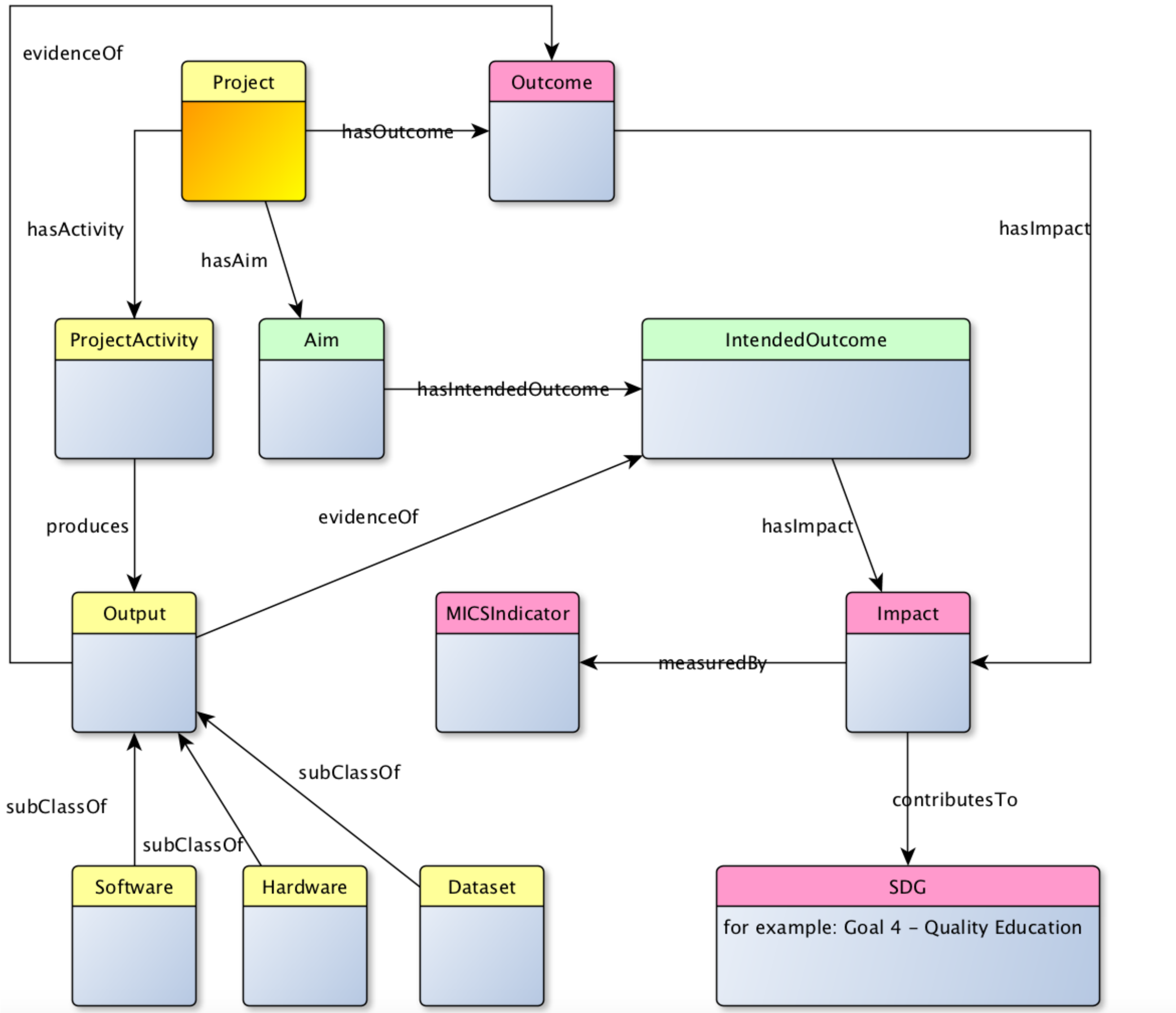
Property	**Description/comments**
hasAim	It associates a project to its aims.
hasKeyword	It associates a project to its keywords. Keywords are chosen from a predefined set of text expressions.
hasIntendedOutcome	It associates the intended outcomes of a project to its aims. For example, a project can have, as an aim, "to evaluate impacts of citizen science" and, as the intended outcome, "a greater understanding of the meaning of impact". As another example, a project can have, as aim, "to increase the environmental impact of citizen science" and, as the intended outcome, "an increase of the awareness of the citizens".
hasOutcome	It associates a project to its outcomes.
hasTag	It associates a project to its tags. Tags are free text expressions. The keyword vocabulary can be used for tags.

Figure 4. The MICS Project Detailed Information classes and properties

3.4 Project Impact

The MICS ontology Project Impact module is the critical metadata scheme relating to the impact a citizen project has, how it is measured, and how it relates to the SDGs described in the introduction of this report. How a projects' impact can be measured will be explained in detail in deliverables D2.2:

Report detailing impact-assessment methods adapted to citizen science and D2.3: Impact-assessment methods adapted to citizen science. It follows on from the “Outcome” and “IntendedOutcome” classes introduced in the Project Detailed Information module, bringing in classes and properties relating to outputs, impact, SDGs and impact assessment indicators. Figure 5 illustrates the Project Impact Module, along with its classes and properties.



Property	**Description/comments**
hasActivity	It associates a project with the activity/s it asks participants to carry out.
produces	It associates the outputs of a project to the activity that produced them.
subClassOf (Software, Hardware & Dataset)	It defines the different categories of outputs a citizen-science project can have.
evidenceOf	It associates an output as evidence towards an outcome, be it intended or not.

hasImpact	It associates a projects' outcomes, and intended outcomes, to the impact they have.
measuredBy	It associates a project impact with the indicator that defined it, for example, "E.1: Natural habitats enhanced".
contributesTo	It associates a projects' impact with the Sustainable Development Goal (SDG) it contributes towards. For example, the impact "Water quality has improved throughout region A, in regards to nitrate parts per million…" could contribute to the SDG "Goal 6: Clean Water and Sanitation" AND "Goal 3: Good Health and Wellbeing".

Figure 5. The MICS Project Impact classes and properties

The Project Impact module of the MICS ontology presents a number of new data types and their relationships. It associates a project's output and outcomes to its broader impact, whether they are related to well-defined aims of the project or are more general outcomes that occurred more organically. In turn, the project impacts are defined as such and measured by the MICS Impact Assessment indicators (as described in deliverables D2.2 and D2.3). Finally, a project's impact can also contribute to one or more of the UN's sustainable development goals. As with its aims and outcomes, a project can have multiple impacts, measured by various MICS indicators, that contribute to various SDGs.

4 MICS Ontology: Multilingual Considerations

For the MICS Ontology to be adapted across the citizen science landscape, primarily by the MICS test and validation site locations, further field geographically, and by other platforms (EU-Citizen.Science, WeObserve.eu etc.), it needs to be easily translatable and contain terms and concepts that have meaning multilingually. The ontology has been designed with this in mind, initially using terms and wording that are universally translatable, with more advanced mechanisms explored when the platform evolves to perhaps include more nuanced and bespoke examples.

4.1 EuroVoc and Google Translate

The MICS ontology and its associated modules, classes and properties use the EuroVoc dictionary as a basis to ensure that the terms used are multilingual, easy to translate and have a well-defined meaning. EuroVoc (2012) is a highly multilingual thesaurus, consisting of over 6,700 hierarchically organised subject domains used by European Institutions and many authorities in the Member States of the European Union (EU), for the classification and retrieval of official documents. It contains terms in 23 EU languages (Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish and Swedish). It involves controlled vocabularies, schemas, ontologies, and data models. The data models include the EU *Common Data Model* (CDM)[6], an ontology able to represent the relationships between resources managed by the EU Publications Office, which is

[6] EU Vocabularies Common Data Model, https://op.europa.eu/en/web/eu-vocabularies/model/-/resource/dataset/cdm, accessed 19th February 2020

described by using RDF(S) and OWL technologies (hence the MICS ontology following the OWL structure also).

Not only does aligning the MICS ontology with the EuroVoc framework ensure that it is easily translatable across the EU, but it also allows for the use of existing tools to automate the process. Freely available, open-source systems exist that automatically classify text and documents into the EuroVoc multilingual thesaurus (Boella et al., 2013), and that automatically assign EuroVoc descriptors and categorisations to new documents and schemes (Steinberger et al., 2013). As the ontology evolves and grows, such automation will assist in ensuring it remains multilingual and applicable across the citizen science domain.

In addition to the bespoke tools that have been created in association with the EuroVoc framework, by taking this approach, the MICS ontology can also take advantage of universal, web-based translation solutions. For example, Google Translate[7] has been shown to, while far from perfect, produce translations that range from understandable to impressive (Groves and Mundt, 2015). However, such an approach will need validation, especially in the case of translating an ontology which may have only one-word descriptions with little associated context. In the case of the MICS project, the options include the use of the case study project coordinators as experts to validate the ontology before it is deployed, perhaps with the support of an automated system (Arcan et al., 2016).

4.2 The Potential of Crowdsourcing for Translation

As previously mentioned, the MICS ontology and the corpus of knowledge associated with it are designed to be an ever-evolving system, learning from the impact of existing and previous citizen science practices and in turn informing future citizen science endeavours. This evolution could pose significant challenges in terms of ensuring the ontology remains backwardly consistent and multilingual. Many of these challenges are perhaps administrative in form, with the MICS project providing systems that align any updates or progress to the ontology to a corresponding update to the translation.

However, there are also potential challenges that are more nuanced. As the MICS ontology and corresponding framework reach out beyond considering the more 'mainstream' forms of citizen science endeavours, there is the possibility of encountering more bespoke, 'one-off' approaches to citizen science. To deal with this variability, the EU-Citizen.Science project is developing a set of characteristics to create a common understanding of what citizen science involves[8]. With such examples, more specific, technical language to describe data and relationships might exist. In such cases, more universal, machine-led examples such as Google Translate can be initially less successful in providing adequate translation (Patil et al., 2014).

In recent times, specifically from the release of Web 2.0 onwards, crowdsourcing has become a widely used solution in the analysis and processing of large datasets. This processing includes translation, where crowdsourced solutions, in the past, have been more accurate when considering more specific, technical language compared to the machine-led alternative (Anastasiou and Gupta, 2011). While such an approach requires a form of 'quality control' to prevent disfluent, low-quality results (Zaidan and Callison-Burch, 2011), and while its performance will likely be surpassed by machines in the near

[7] Google Translate, https://translate.google.co.uk/, accessed 19th February 2020
[8] EU-Citizen.Science, http://eu-citizen.science/what-is-citizen-science/, accessed 16th March 2020

future, existing citizen-science platforms have already adopted it with some success. The Zooniverse first started using crowdsourced translation in 2014, and have successfully translated their online citizen science projects into numerous European languages, and some of its projects even further afield into Traditional Chinese[9]. In fact, due to the success of this enterprise, the approach has been adopted in the opposite direction, being used as a tool to educate and train future translators with a ready to use platform that can be directly compared across several languages (Michalak, 2015).

As such, the crowdsourcing of translation is a potential mechanism for the MICS project to consider to ensure the ontology and framework remains multilingual, even as its reach includes specific terms and bespoke language that is perhaps beyond the scope of the EuroVoc dictionaries. Crowdsourcing and general machine-led solutions will allow the MICS framework to reach beyond the scope of the members of the consortium, specifically the validation and test case studies locations, and into areas perhaps less traditionally engaged by the citizen science domain.

5 References

- Anastasiou, D. and Gupta, R. (2011). Comparison of crowdsourcing translation with Machine Translation. *Journal of Information Science*, 37(6), 637–659.
- Arcan, M., Dragoni, M. and Buitelaar, P. (2016). Translating Ontologies in Rea-World Settings. *The Semantic Web – ISWC 2016,* 241-256.
- Boella, G., Di Caro, L., Rispoli, D. and Robaldo, L. (2013). A system for classifying multi-label text into EuroVoc. *ICAIL '13, Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law,* 239-240.
- Groves, M. and Mundt, K. (2015). Friend or Foe? Google Translate in language for academic purposes. *English for Specific Purposes,* 37, 112-121.
- Michalak, K. (2015). Online Localisation of Zooniverse Citizen Science Projects – On the Use of Translation Platforms as Tools for Translator Education. *The Journal of Teaching English with Technology,* 3, 61-70.
- Patil, S. and Davies, P. (2014). Use of Google Translate in Medical Communication: Evaluation of Accuracy. *BMJ,* 2014:349, 7392.
- Steinberger, R., Ebrahim, M. and Turchi, M. (2013). JRC EuroVoc Indexer JEX – A freely available multi-label categorisation tool. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC '2012),* 798-805, Istanbul, May 2012.
- Zaidan, O. F., Callison-Burch, C. (2011). Crowdsourcing translation: professional quality from non-professionals. *HLT' 11: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies,* 1(1), 1220-1229.

6 Appendix

The OWL code of COST Action CA15212 Working Group 5's Citizen Science Ontology, adopted by MICS, can be found at [https://zenodo.org/record/3674917#.Xk-sTGj7S70].

[9] The Zooniverse Blog: Zooniverse Translations Update, https://blog.zooniverse.org/tag/translation/, accessed 21st February 2020