Downloaded from orbit.dtu.dk on: Apr. 12, 2023

DTU Library

# Research data management challenges in citizen science projects and recommendations for library support services. A scoping review and case study

**Hansen, Jitka Stilund; Gadegaard, Signe; Hansen, Karsten Kryger; Larsen, Asger Væring; Møller, Søren; Thomsen, Gertrud Stougård; Holmstrand, Katrine Flindt**

*Published in:*
Data Science Journal

*Link to article, DOI:*
10.5334/dsj-2021-025

*Publication date:*
2021

*Document Version*
Peer reviewed version

Link back to DTU Orbit

*Citation (APA):*
Hansen, J. S., Gadegaard, S., Hansen, K. K., Larsen, A. V., Møller, S., Thomsen, G. S., & Holmstrand, K. F. (2021). Research data management challenges in citizen science projects and recommendations for library support services. A scoping review and case study. *Data Science Journal*, *20*(1), 1-29. [25]. https://doi.org/10.5334/dsj-2021-025

**General rights**
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Title:** Research data management challenges in citizen science projects and recommendations for library support services. A scoping review and case study.

**Authors:** Jitka Stilund Hansen[1], Signe Gadegaard[1], Karsten Kryger Hansen[2], Asger Væring Larsen[3], Søren Møller[4], Gertrud Stougård Thomsen[5], Katrine Flindt Holmstrand[1]

**Affiliations:** [1]DTU Library, Technical University of Denmark, [2]Library, Aalborg University, [3]University Library of Southern Denmark, [4]Roskilde University Library, [5]Aarhus University Library, Royal Danish Library, Denmark

**Corresponding author:**

Jitka Stilund Hansen, DTU Library, Technical University of Denmark

E-mail: jstha@dtu.dk

**Authors' contributions:**

JSH, SG, KFH, AVL designed and did the literature searches. All authors participated in screening of the retrieved publications and JSH extracted data from included publications. KFH, JSH, GST, SM, AVL did the interviews and extracted data for the case study. JSH drafted the manuscript and all authors commented and approved it.

# Abstract

Citizen science (CS) projects are part of a new era of data aggregation and harmonisation that facilitates interconnections between different datasets. Increasing the value and reuse of CS data has received growing attention with the appearance of the FAIR principles and systematic RDM practises, which are often promoted by university libraries. However, research data management (RDM) initiatives in CS appear diversified and if CS have special needs in terms of RDM is unclear. Therefore, the aim of this article is firstly to identify RDM challenges for CS projects and secondly, to discuss how university libraries may support any such challenges.

A scoping review and a case study of Danish CS projects were performed to identify RDM challenges. 48 articles were selected for data extraction. Four academic project leaders were interviewed about RDM practices in their CS projects.

Challenges and recommendations identified in the review and case study are often not specific for CS. However, finding CS data, engaging specific populations, attributing volunteers and handling sensitive data including health data are some of the challenges requiring special attention by CS project managers. Scientific requirements or national practices do not always encompass the nature of CS projects.

Based on the identified challenges, it is recommended that university libraries focus their services on 1) identifying legal and ethical issues that the project managers should be aware of in their projects, 2) elaborating these issues in a Terms of Participation that also specifies data handling and sharing to the citizen scientist, and 3) motivating the project manager to good data handling practises. Adhering to the FAIR principles and good RDM practices in CS projects will continuously secure contextualisation and data quality. High data quality increases the value and reuse of the data and, therefore, the empowerment of the citizen scientists.


Keywords: research data management, citizen science, FAIR principles, university library

# Introduction

The citizen science (CS) method has broad perspectives in using citizen-driven data collection to answer research questions and address societal challenges in all fields of science. From a scientific perspective, involving interested members of the public in the generation of large, spatially and temporally highly complex data sets is one of the greatest benefits of CS. CS projects are often initiated as a collaboration between scientists and lay people, but initiatives driven by non-academic individuals, communities or private organisations are widespread globally.

With the availability of new easy-to-use technologies, data collection by the volunteers increases in volume and sophistication. Already, CS projects are part of a new era of data aggregation and harmonisation that facilitates interconnections between different datasets. Therefore, CS data have the potential to form the foundation of innovations, new discoveries and policymaking.

The European Citizen Science Association has developed Ten Principles of Citizen Science Projects that defines its view of good practices in CS (ECSA, 2015). Among these, is the encouragement to make project data and metadata publicly available and if possible publish results in open access format (Principle no. 7). Apart from being of benefit to both the professional and the citizen scientist (Principle no. 3), CS is generally viewed as having a communal output through data sharing and openness. For example, CS is one of the eight pillars of Open Science identified by the Open Science Policy Platform, an EC Working Group (OSPP, 2017).

In order to create data that are open and meaningful to the community, management of the data has to be considered throughout the data life cycle. Thus, research data management (RDM) encompass measures to ensure the usability and reusability of research data before, during and after the research project (Holmstrand et al, 2019). The FAIR guiding principles for research data can be used for this work and for generating future-proof and machine-readable data (Wilkinson et al, 2016).

In 2016, a survey from the Joint Research Centre (JRC) found RDM practises in CS fragmented and although the respondents wished to share the project data, apps and services, their interoperability and reusability were not secured (Schade and Tsinaraki, 2016). A recent study found that in general, CS projects were not implementing or being aware of best practices for RDM (Bowser et al, 2020). However, international and national RDM initiatives emerge and reflect a growing attention to ensuring consistent RDM.

RDM as a structured discipline and gathering concept is still a rather new area where a multifaceted skill set is needed, often one beyond the scientific focus. At the university, joint RDM activities are largely embraced and developed by the library for example by offering repositories and data curation, metadata and information system specialisations (Corrall, Kennan and Afzal, 2013; Karasmanis and Murphy, 2014). Increasing demands for sharing research data openly or securing their reusability and the national and international endorsement of the FAIR principles, have given the university libraries the opportunity to advocate for, support and train in FAIR data and RDM.

In 2019, a Danish project was launched to investigate the possibility of libraries to promote and support the propagation of CS. A part of this project was to identify where university libraries could focus their services towards the CS discipline and naturally, the consideration of RDM services were included. However, if CS would have special needs in terms of RDM were not clear. Therefore, the aim of this article is firstly to identify RDM challenges for CS projects and secondly, to discuss how university libraries may support any such challenges. Summary of the identified challenges are provided in the last section as basis for the recommendations for the university libraries guiding CS project managers.

# Methods

To identify RDM challenges for CS projects, we conducted two studies; A scoping review retrieving reviews, book chapters, reports, articles and internet resources and a case study of four Danish CS projects consisting of interviews with the principal investigator. By conducting a scoping review with a systematic literature search, we aimed to advance our knowledge of the current state of RDM in CS and identify key themes on which to focus library practices. The case study was conducted with the same intentions and to confirm if the findings of the literature study were representative of challenges in Danish academia-based CS projects.

## Scoping review strategy

Two questions formed the base of a systematic literature search: 1) What challenges are CS projects facing in terms of RDM? 2) Are the FAIR principles applied for data in CS projects?

Appendix 1 (Supporting Text 1) shows the systematic literature search performed in Scopus and Web of Science to answer these questions. The search focused on legal and ethical aspects, intellectual property rights (IPR), as well as issues related to sharing and reuse of data. A broader Google search and a search in BASE (Bielefeld University Library, n.d.) was also done. Appendix 2 (Supporting Text 1) describes the screening process, the eligibility criteria and contains a PRISMA diagram (Moher et al, 2009) of the process.

## Data extraction from the publications

We summarised the included publications descriptively and inferred the RDM challenges if not directly described. Table 1 categorises content into findability, accessibility, interoperability, reusability (FAIR) and general aspects of research data management and related infrastructures. Table 2 presents publications concerned with ethical and legal issues. Some publications state recommendations or solutions to the problems presented, which are also included in the data extraction. Table 3 is a collection of published tools, guidelines and formal recommendations, which directly encompass issues related to RDM in CS projects. We did not search specifically for publications describing guidelines and recommendations, but have included and categorised them, because of their relevance to our investigation.

## Case study

Four Danish CS projects were included as cases and identified through the authors' universities. One project has a health focus and the remaining are focused on biodiversity in Danish waters or litter in the Danish terrestrial environment. Semi-structured interviews (Appendix 3, Supporting Text 1) were performed with the leading scientists of the projects, who are all university employees. They were asked about the project data flow, their knowledge of the FAIR principles and RDM issues in their projects. Table 4 describes the projects and data are extracted to Table 5 with the same foci as Table 1 and 2.

## Limitations

We performed a comprehensive search with the specific focus on "citizen science". One limitation of this study may be that words such as "crowd-sourcing" or "volunteer monitoring" were not used and could have omitted useful references. However, our search did retrieve references associated with comparable initiatives such as crowd-sourcing and other participatory research. Taking into account the differing use of the term "citizen science", we obtained a broad range of references, deeming the review methodology appropriate. Because we did not search specifically for guidelines and tools, the search may not be exhaustive. Other guides and tools for CS projects may have been excluded because aspects of RDM were not addressed.

Our case study is very small and only encompasses professional scientists performing CS projects. Also, the cases are only Danish, which may represent a rather geographically restricted group regarding adherence to national and institutional policies, but also regarding level of institutional RDM services and knowledge of the FAIR principles. Last, all authors are affiliated with university libraries which may bias our focus towards supporting CS arising from academia.

# Results and discussion

## RDM challenges identified from literature search

### Knowledge of and adherence to the FAIR principles

The selection criteria of this review generally excluded individual CS projects, so how widespread the practical implementation of the FAIR principles is cannot be determined. Of the 48 included articles, only three directly mention and work with the FAIR principles

(Bastin, Schade and Schill, 2017; Clements et al, 2017; Kissling et al, 2018). One of these articles addresses Volunteered Geographic Information (VGI), the two others are summaries of working group (WG) meetings within air sensor monitoring and Essential Biological Variables. Furthermore, among the identified guidelines and tools (Table 3), the DM system developed by Ocean Network Canada adheres to the FAIR principles (Wolf et al, 2019). The two WG summaries and the ONC system are not only directed towards CS data, indicating that the FAIR principles could find its way to CS through international organisations and communities embracing CS. However, most of the included articles and guidelines address RDM challenges (and their solutions), which are encompassed in the FAIR principles, hence the data presentation in Table 1 is shaped accordingly.

## Findability

The ability to discover data, the findability aspect of the FAIR principles, is only indirectly or not at all addressed in most of the included articles. For instance, natural history collections may provide data for CS projects. However, Runnel and Wijers (2019) describe that it is currently not possible to search for natural history collection data in CS portals. i.e websites where CS projects are displayed or where CS data are published. With offset in the PPSR-CORE Program Data Model Metadata Standard (US CSA Data and Metadata WG, 2019), they suggest which metadata fields may accommodate the need for storing and finding information about natural history collections that form the basis of CS projects.

Therefore, one challenge for CS project data management is to make data findable and also identified as of CS origin. This leads to the associated challenge that platforms to accommodate CS data or discipline specific data could be used more systematically by CS project managers to increase the discoverability and reuse of data.

Adriaens *et al*. (2015) recommend the Global Biodiversity Information Facility (GBIF) as a publishing platform for CS project data on invasive species, because of the use of metadata standards and the possibility to share and not the least find such datasets. If existing platforms can provide alerts to stakeholders monitoring and handling invasive species, this could create an automated system for finding the newest data.

According to the FAIR principles, data must be assigned a persistent identifier (PID), such as a DOI, for permanent findability. A general challenge for evolving datasets, such as many CS data, is how to cite and retrieve a subset of a dataset as it existed at a specific date and time (August et al, 2015; Hunter and Hsu, 2015). The Research Data Alliance (RDA) Data Citation WG has developed a Recommendation based on two principles (Rauber et al, 2015): first, one must ensure that data are stored in a versioned and timestamped manner; second, the PID to the citable data should comprise a query to the dataset and a timestamp. Hunter and Hsu (2015) found the principles highly applicable to a test CS dataset.

## Accessibility

Citizen scientists often engage in projects because of personal interests and expertise. Such interests can be based on leisure activity interests (bird watching), but also based on engagement in issues that affect the environment or well-being of a community (Ganzevoort et al, 2017; Kennan, Williamson and Johanson, 2012). Crall *et al*. (2010) found that volunteers expected access to data and they deemed it more important to readily share data than waiting to release data until after scientific publication of results. This is in line with the general view of CS as a discipline, where data is shared at large. August et al. (2015) states that access must also be secured by good data curation. Further, keeping data accessible may promote data quality control and reuse (Kissling et al, 2018). Academic researchers may be reluctant to share data before they have published their findings, however, moving from data sharing (i.e. providing access under specified circumstances) to data publication with the possibility to get cited may be a motivation to make data open access (August et al, 2015; Groom, Weatherdon and Geijzendorffer, 2017). Also, a study from JRC found a great interest among CS project leaders to provide access to the data, but this was not reflected in what was actually being done (Schade, Tsinaraki and Roglia, 2017; Schade and Tsinaraki, 2016).

Therefore, the challenge of many CS projects is how to accommodate the wish for data access to the volunteers or the public, including the scientific community. This should be weighed against the other challenge of changing the incentives for academic researchers to publish data and therefore, promote the reuse of their data.

If and how data can be accessed may largely rely on the content of private or sensitive information embedded in the data. Several articles of Table 1 and 2 investigate the challenges of handling such information and propose strategies for balancing it. The most evident challenge of many CS projects is how to protect the personal information (name, contact information etc.) of the volunteers and how to handle their location sharing. Also, collecting data on private land could indirectly expose land ownership. Furthermore, security for objects collected must be considered, e.g. location of endangered species or unintentional photo capture of persons or secondary objects (Anhalt-Depies et al, 2019; Bowser et al, 2014; Groom, Weatherdon and Geijzendorffer, 2017; Higgins et al, 2016; Williams et al, 2018). Lastly, observations may contain sensitive information about a people or region that they may not want to share openly (Pulsifer, Huntington and Pecl, 2014).

A survey of CS projects of invasive species found that these concerns pose very practical threats in terms of data access (Crall et al, 2010) and without support on how to navigate, this would be a reason for project managers not to share CS data openly. Interestingly, citizens engaged in CS often focus on sharing and openness for common benefits, and evaluate their own privacy concerns in the context of the project (Bowser et al, 2017). Several articles put forward recommendations (Anhalt-Depies et al, 2019; Bowser et al, 2017, 2014; Resnik, Elliot and Miller, 2015; Williams et al, 2018) that can be summarised as: i) collect as few personal and sensitive data as necessary, ii) obfuscate such information upon publication or sharing and iii) clearly inform the participants of what will be shared, why it is necessary and how it will be done. Refer to Table 2 for an elaboration and see the section below on protection of private data.

## Interoperability

The quality of CS data is closely interlinked with how the data are described and with what content (metadata and other documentation) data are published. Describing data with rich metadata and using metadata that follow specific standards or community-recognised ontologies is important for securing interoperability (GO FAIR, n.d.). One example is from the air monitoring sensor workshop document (Clements et al, 2017). Low-cost air quality sensors are widely used and important for empowering communities. However, their deployment has not been followed by standards for data formats, units and for metadata and therefore, exchange of data between communities is often not possible without data transformation or excessive processing. The same conclusion is reached for new technologies developed to study the biological world (August et al, 2015) and for VGI data (e.g. websites, apps, instant species and location definition)(Bastin, Schade and Schill, 2017). Thus, data that are not interoperable have very low value in the perspective of the general public (community interoperability)(Williams et al, 2018) or regulatory authorities (Owen and Parker, 2018). Results from scrutinized biomedical CS platforms (Borda, Gray and Fu, 2020) and a CS project survey (Schade and Tsinaraki, 2016) revealed that use of standardised data and metadata was not supported or rarely used, respectively. Whether this is because appropriate standards are unavailable or difficult to use, is unknown. Thus, the next RDM challenges identified for CS is supporting and creating interoperable data of quality and value, supported by accessible standards, and that ventures in new technologies should follow community standards.

One important step towards solving this challenge is performed by the CS COST Action and several international partners, who aim to extend a standard on key elements and concepts of CS (De Pourcq and Ceccaroni, 2018) based on the existing PPSR-Core (US CSA Data and Metadata WG, 2019). The ontology encompasses a project metadata model, a dataset metadata model and an observation data model. The ontology is based on existing standards; the Open Geospatial Consortium standards, ISO/TC 211, W3C standards (semantic sensor network/Linked Data), and existing GEO/GEOSS semantic interoperability (COST Action CA 15212, 2019). Guidelines for its implementation and retrofitting into existing platforms will be provided in the future.

Publishing primary biodiversity data is often done with the Darwin Core Standard and Access to Biological Collection Data. The Ecology Metadata Language is widely used for the ecology discipline and all are used or adapted by the data aggregator GBIF. These standards not only ensure semantic interoperability between datasets and disciplines, but also machine-readability. Both semantic interoperability and machine-readability are called for in several articles, again underscoring that this ensures the long-term use and secures the data against technological changes (August et al, 2015; Bastin, Schade and Schill, 2017; Kissling et al, 2018; Simonis, 2018; Williams et al, 2018).

## Reusability

Access to data can be meaningless if data are incomprehensible or difficult to extract. For a volunteer, aggregated and processed data may be more relevant than for a scientist or governmental authority in need of raw data. In both instances, data lose their value without explanation of the provenance or context (Sheppard, Wiggins and Terveen, 2014; Williams et al, 2018). The review by Borda, Gray and Fu (2020) revealed that documentation of data provenance or context across the data life cycle varies largely on biomedical CS platforms. Policy-making bodies, such as environmental protection agencies, can only use data of certain quality (Owen and Parker, 2018) and the same applies for CS data incorporated in scientific publications (Williams et al, 2018). How to obtain and support good quality CS data is not addressed in this review, but it is inevitably linked to the possibility of reusing the data. Therefore, the challenge for CS projects in order to promote the reuse and secure the long-term value of collected data is to document why and how data were collected, if changes in sampling protocols occurred, and how data were processed. This documentation should follow the data, possibly by integration in the metadata.

Another challenge of CS projects related to reuse of data is the lacking application of data licenses. The GBIF is a platform for sharing biodiversity data and a survey into use of data licenses revealed that only 3% of CS datasets had a data license (Groom, Weatherdon and Geijzendorffer, 2017). It is generally perceived that not applying a license severely hampers the open use of data (Groom, Weatherdon and Geijzendorffer, 2017; Williams et al, 2018). Also, the JRC survey on practices in CS projects revealed that data licensing often is not considered until late in the project, which may cause confusion between volunteers and project management (Schade and Tsinaraki, 2016). Data aggregation is widely used in biodiversity research, why Kissling *et al*. (2018) state that legal interoperability is necessary. Automated workflows during aggregation of different datasets are facilitated if the used licenses are interoperable. For example, the use of an aggregated dataset will be restricted if the two underlying datasets are CC BY-ND and CC BY, respectively (Kissling et al, 2018).

Some CS projects allow upload of images or media files as part of the data collection. However, if media files do not have a license, then the linking to and use of accompanying data is hampered (Adriaens et al, 2015).

The recommendations from the included articles can be summarised: (i) organisations must implement clear licensing policies and projects could make the volunteers choose license for their own data (Groom, Weatherdon and Geijzendorffer, 2017), (ii) inform users about issues of IPR of records and associated media files so that this does not restrict further usage (Adriaens et al, 2015), and (iii) use CC0 and CC BY to promote legal interoperability (Kissling et al, 2018). Further, making the volunteers choose a license for the data they collect will require automated processes for data extraction and should be aligned to ease legal interoperability.

## General research data management and infrastructures

Many CS projects and research areas suffer from the lack of available infrastructure such as tools for collecting data, databases, publishing platforms i.e. data management systems (August et al, 2015; Clements et al, 2017; Crall et al, 2010). The conclusions from the workshop on air quality measurements was that the community would hugely benefit from a large-scale data management system that could offer interoperable and shareable data for comparisons (Clements et al, 2017). The Global Invasive Species Information Network aims to link online data sources on invasive species and finds that CitSci.org may accommodate CS projects' data and privacy concerns and their need for publishing data (Crall et al, 2010). Where GBIF could be a tool for sharing invasive species data with the scientific communities and authorities (Adriaens et al, 2015), CitSci.org is developed for project and data management of CS projects in general, offering use of existing metadata standards for quality assurance and interoperability (Wang et al, 2015).

However, in order to increase the ability to access and reuse of for example environmental data, there is a need for infrastructures to be developed and provided for by authorities, such as environmental protection agencies (Owen and Parker, 2018), or, which already occurs, by consortia funded for example by the EU (Higgins et al, 2016).

Access to DM systems and infrastructure may be another very practical challenge for remote communities such as those of the Arctic (Pulsifer, Huntington and Pecl, 2014). RDM is not always only about technical solutions, but should be fitted to reflect local culture and economy. However, securing a locally embedded DM system will support knowledge exchange not only for the scientists but for the communities as well (Pulsifer, Huntington and Pecl, 2014). Chimbari's experiences with data collection in South Africa makes him stress that clear DM policies and agreements on how data is returned from data collector to the principal investigator are necessary to secure the data (Chimbari, 2017).

Another RDM challenge of CS is how to sustain interoperability of software or technology used in CS projects (Adriaens et al, 2015). This is addressed by the Air Sensor Workgroup that works to make software, technologies and data platforms in open source so users can implement and further develop the tools to their needs (Clements et al, 2017). However, many projects develop apps and platforms that are never reused because of discontinuation of the project or unavailable documentation.

However, to save and share resources, project resources must be allocated to RDM. This challenge is well known, since many projects can't guarantee sustained or any access to data – either because of lack of skills, insufficient funding (Schade and Tsinaraki, 2016) or simply because it has not been considered spending resources on (Adriaens et al, 2015). Based on the widespread occurrence of projects that collect data on invasive species, Adriaens *et al.* (2015) stress that sustainable funding is much needed to secure data and technological support in the long-term. A call for funders to recognise that access to quality data requires committed funding (Bastin, Schade and Schill, 2017) is now accommodated by the Horizon 2020 Open

Research Data Pilot, where funding can be allocated to data management and securing open access to data (European Commission, n.d.).

## Authorship and recognition of citizens

One of ECSA's 10 principles states; "Citizen scientists are acknowledged in project results and publication". However, there is no consensus on how this is done (Tauginienė, 2019). Accordingly, several of the publications in Table 1 and 2 address the challenges associated with recognition of volunteers and with co-authorship for citizens on scientific publications. Currently, scientific journals follow the ICMJE criteria for authorship (ICMJE, n.d.), which exclude citizens to be attributed co-authorship (Resnik, Elliot and Miller, 2015; Ward-Fear et al, 2020). Authorship or formal recognition is, however, an important tool to give back something to volunteers, but also to prevent their exploitation (Resnik, Elliot and Miller, 2015).

Ward-Fear *et al*. (2020) propose the implementation of group co-authorship to cohorts of non-professional scientists. The authors use the example of the Balanggarra Rangers, who were included as group co-authors on two scientific publications on an Australian conservation intervention. The intervention could not have taken place without the Rangers' knowledge as traditional owners of the land and their huge involvement in the study. Because of the obstacles with giving authorship to a large number of individuals (Ward-Fear et al. 2020), recognitions can also be performed in the acknowledgement section of a paper (Resnik, Elliot and Miller, 2015). Groom, Weatherdon and Geijzendorffer (2017) argue that recognition of contribution from citizen scientists should be supported by the data users, if citizen scientists for example may wish for a recognition of the work performed in their community. Another solution was explored by Hunter and Hsu (2015), who were able to credit individual citizen scientists contributing to a specific data subset. They based their initiative on RDA's Dynamic Data Citation approach (Rauber et al, 2015). Interestingly, ca. 40% of biodiversity volunteers would like to be cited by name, when their data are used (Ganzevoort et al, 2017).

## Intellectual property rights

Williams *et al*. (2018) allocate IPR considerations to two entities: (i) "background IPR" that encompasses how knowledge and data will be used and under what restrictions and (ii) "foreground IPR" that should consider how the project allows access to the knowledge and data. This paragraph is concerned with the challenges of background IPR in CS projects, while foreground IPR was discussed in a previous section under "Accessibility".

Through their engagement in CS projects, citizens may develop photographs, writings, and creative selections or arrangements of scientific data (Guerrini et al, 2018). Such creations could cause IPR disagreements. In contrast to the undisputable regulations in many countries of employees' inventions, volunteers in CS retain the IPR to any copyrightable work they produce. Therefore, patent assignment cannot readily be performed by a principal investigator, because citizens possess the right to exclude the CS project in using a CS invention they have produced (Guerrini et al, 2018). Another more ethical question surrounds the sharing of culturally embedded knowledge. Traditional knowledge should be treated with respect, in particular if communities expect to retain some control over gathered data (Resnik, Elliot and Miller, 2015).

General recommendations (Table 2) are to make transparent IPR agreements that are regularly updated with the volunteers (Guerrini et al, 2018; Williams et al, 2018) and that the scientist (or project holder) should aim at sharing IPR, education or monetary value with the volunteers (Resnik, Elliot and Miller, 2015). Also, refer to the section above on licensing and legal interoperability (Reuse of data).

## Participant protection and privacy

Laws and policies protect participants of scientific studies, and studies involving human subjects will under many circumstances require ethical permission by a national, regional or institutional ethical committee (EC). The aim of the EC review is to protect subjects from harm, and oversee inclusion and exclusion criteria as well as recruitment and informed consent procedures. In addition, the risk of vulnerable populations' participation and the procedures to cope with incidental findings are evaluated.

Several articles in Table 2 originate from the US where the Common Rule is a federal policy to protect human subjects in research, where biospecimens or identifiable data are collected. The Common Rule regulates all government-funded research and virtually all American academic and health care institutions adhere to it independent of their funding and use it during institutional review board (IRB) reviews (Rothstein, Wilbanks and Brothers, 2015). However, in some contexts CS participants are not regarded as research subjects, but rather as "research assistants" and the Common Rule does not mandate IRBs to consider risks or benefits to citizens who facilitate research in other ways (Guerrini et al, 2018; Oberle et al, 2019; Rothstein, Wilbanks and Brothers, 2015). Also, another challenge that the authors describe is that private initiatives such as community-driven CS projects fall outside the Common Rule and do not have to go through IRB review (Guerrini et al, 2018; Patrick-Lake and Goldsack, 2019; Wiggins and Wilbanks, 2019).

Biomedical research is a primary example of an area where this challenge is evident. The current technology provides us with apps and gadgets collecting personal health data, which individuals may choose to donate to projects not subjected to academic regulation and policies. In some cases, participants may not be able to fully understand how and by whom their data are used, because of obscured content of the informed consent (Patrick-Lake and Goldsack, 2019; Rothstein, Wilbanks and Brothers, 2015; Wiggins and Wilbanks, 2019). The collection and aggregation of health data could reveal health issues causing distress to the participant. In clinical research, the disclosure of incidental findings is regulated by policies and performed by clinicians, but in CS, these findings may either not be disclosed to the participant or the participant may be left alone with the observations (Guerrini et al, 2018; Rothstein, Wilbanks and Brothers, 2015).

Some CS researchers may wish for legal guidance and EC or IRB review, which may not be a possibility within the current ethical frameworks unless funding for this is obtained (Guerrini et al, 2018; Wiggins and Wilbanks, 2019). Therefore, it may be necessary to clarifying ethical issues for example in a national ethical framework for CS (Bonn et al, 2016) or by extending existing policies (Guerrini et al, 2018).

These challenges may be relevant for CS projects in countries, where CS projects fall outside national laws and academic policies. In Denmark, all research with human subjects, where biological specimens are collected or biological processes recorded during an intervention, is regulated by the Act on Research Ethics Review of Health Research Projects (Danish Parliament, 2011), which may guide CS projects both of academic and non-academic origin.

In the EU, the GDPR regulates the protection of data and privacy, and applies to all handling of personal data by businesses and organisations; this refers to data that can identify a person, but also sensitive data such as information on health, ethnicity, religion etc. Not all states of the USA have laws protecting privacy or sensitive information of participants in for example CS projects. Therefore, many data handlers will not be obliged to protect data or inform participants on security breaches and they can give or sell access to data to third-parties (Rothstein, Wilbanks and Brothers, 2015).

Another legal question is that insurance coverage conditions often are unclear, when doing research including volunteers. This is in contrast to research subjects, who for example in Denmark are covered by the public patient or work injury insurances (NVK, 2017) Therefore, a German green paper recommends setting up extended insurance for volunteers actively participating in CS projects (Bonn 2016).

Overall, the challenge for many CS researchers is how to balance the assets of open science and the engagement and trust of the participants with ethical and legal obligations, in particular if no clear framework exists for the latter.

## Research integrity

Another ethical concern is that direct publication of non-academic CS data without peer-review and/or quality control can lead to misinformation (Wiggins and Wilbanks, 2019). On the other hand, the need to assess validity and facilitate discussion of the results may not be fulfilled, since private CS projects are not obliged to share or publish data (Rothstein, Wilbanks and Brothers, 2015). Data sharing with participants constitutes one of the principles of CS (ECSA, 2015) and allows the participants and others to reuse, discuss and give feedback (Resnik, Elliot and Miller, 2015).

Finally, disclosing the origin of project funding and of conflicts of interest are necessary to secure transparency and inform about the context in which data were collected (Guerrini et al, 2018; Resnik, Elliot and Miller, 2015; Riesch and Potter, 2014). These publications state this as vital information for others wishing to reuse the collected data (Table 2).

### Existing tools and guidelines

Table 3 is an overview of identified tools and guidelines directed at RDM of CS projects. The references also highlight the challenges described above and/or provide recommendations for RDM. Several identified platforms are directed at CS projects (Bonn et al, 2016; Disney et al, 2017; Greshake Tzovaras et al, 2019; Heigl et al, 2018; Wang et al, 2015) or are scientific project platforms that also can accommodate CS projects (Wolf et al, 2019). The possibilities for handling RDM aspects on these platforms vary widely from simply being a place to store and share data (Anecdata.org (Disney et al, 2017)) to the Ocean Network Canada that provides a complete system for RDM that simultaneously FAIRifies data (Wolf et al, 2019).

Two comprehensive tools for handling RDM issues throughout the data life cycle were identified; one from a DataOne WG (Wiggins et al, 2013) and one from the US Environmental Protection Agency (US EPA, 2019). They also provide step-by-step guidance or templates to writing a data management plan (DMP). A workshop developed principles for using mobile apps and platforms in CS projects and these principles are clearly applicable to the RDM of CS projects in general (Sturm et al, 2018). Several other handbooks and recommendations for CS projects were also identified (Table 3) that stressed the importance of good data handling and/or emphasized the need to resolve any legal constraint on collecting and using data (Pettibone et al, 2016; Tweddle et al, 2012; UKEOF's Advisory Group, 2013; US EPA, 2019; US GSA). An article published after our literature search is also a good source for recommendations aimed at RDM challenges and practices in CS (Bowser et al, 2020).

In 2016, a green paper analysed the requirements and potential of CS initiatives in Germany (Bonn et al, 2016). The following road map recommendations were concerned with the establishment of infrastructures for supporting data management of CS projects, but also providing legal, ethical and collaborative frameworks to support the challenges within these areas. This work is continued in the network platform *Bürger schaffen Wissen*.(Bürger schaffen Wissen, n.d.). The CS Network Austria has established a comparable CS project platform *Österreich forscht* (CSNA, n.d.). In order to use and list your project on the platform, a range of quality criteria have to be met by the user, such as sharing data openly when possible, establishing a DMP and clearly describing ethical and legal data governance (Heigl et al, 2018). The CS Network Austria provides feedback and support in order for the users to meet the listing criteria.

## RDM challenges identified in Danish CS projects

None of the included cases had developed a formal DMP or were aware of the FAIR principles (Table 5). A major obstacle for adopting the FAIR principles for project data and for doing systematic RDM is the lack of time and resources within the project; it has not yet become common practice to include funding for RDM in project proposals and budgets and it is generally not required by funding agencies. Further, RDM support services at the universities hosting the CS projects either do not exist or have been overlooked by the researchers. However, the project leaders expressed interest in using the services more systematically.

The project, Fyn finder marsvin, from 2019 collects a simple dataset that is available via the project webpage and in Zenodo (Table 5). Fangstjournalen aggregates collected data and publishes them regularly on Facebook as a clear strategy to sustain the anglers' motivation to be involved and show the data being utilised. The schoolchildren collecting plastic litter (Masseeksperimentet) can use their own datasets in the class teaching and the data were submitted with a publication and is now available. This underscores that the projects want to share their data or parts of them. Because of the current academic reward systems, the project leaders generally perceive full open access to the data as incompatible with their need to exploit the dataset fully and publish scientific articles before data are released (Table 5). However, one is interested in publishing descriptive metadata of the project in a repository for increasing findability, when presented with the idea.

The projects have not focussed on producing interoperable data defined as including metadata, following standards or ontologies, or data and metadata being described by unique and stable URLs. In general, standardisation is important for the project leaders and one has published a suggestion for standard data to be collected in comparable projects (Venturelli et al, 2017).

Three of the projects contain personal identifiable or location data and the published datasets have removed all personal identification data. When initiated, the dementia projects will contain personal data that cannot be published. One project leader expresses concern about "doing something wrong" if sharing data, because legal counsel is not readily available. The latter, too, is a major barrier for providing access to CS data.

# Knowledge application in the university library

The role of university libraries has evolved with the emergence of new technologies and need for new services (Cox and Corrall, 2013; Karasmanis and Murphy, 2014) and at many universities, the common service surrounding RDM is now founded in the library. Further, the European Commission Open Science Policy Platform WG recommends university libraries as platforms for promoting CS resources and infrastructure (CS WG OSPP, 2018). This review clearly demonstrates that management of CS data faces challenges alike those of other research projects, and therefore supports that university libraries may build on existing resources to become points-of-contact for CS projects.

Several of the identified challenges for CS projects are well known from other research projects and a recent study concluded that CS RDM practices are similar to or lag behind conventional science (Bowser et al, 2020). This means that the university library readily may assist in identifying platforms for setting up and handling CS projects, in using repositories and associated services for data publication, and may guide in the use of appropriate data and metadata standards for the project to secure interoperability. Our findings clearly indicates that applying RDM considerations to the data life cycle will improve the quality and reusability of any CS project and our case study showed that scientists would willingly take the help, which libraries may offer. Therefore, a vital step for libraries with existing RDM support service is to communicate to researchers and CS networks that this expertise already exists.

From the literature and case study, we suggest three focus areas within which the university library could develop more targeted services and recommendations for CS projects; the legal and ethical framework, participant information/contracts and the incentives for allocating resources to RDM.

## Legal and ethical framework for CS data

Several legal issues are part of RDM considerations; however, the library can rarely give legal counsel. The library may therefore support the scientist in identifying and focussing on what legal issues need to be handled and refer the researchers to the institutional legal office.

CS projects often contain personal identifiable information, which requires secure storage and may challenge the CS principle of data being shared openly. An academic project leader should follow the regulation applying to handling of personal data in other scientific projects, but exemplified by our cases, the practical implementation may be confusing and require specific advice.

Fangstjournalen provides a good example on how to balance privacy and participation; the anglers can choose to display their catches or not, and if the data should be part of aggregated data available in the app. However, the scientist can still use the data for research.

The project managers need to be made aware that copyright and IPR can pose constraints on the use of collected data depending on the type of data or knowledge generated. This may affect how to license the data. Further, when CS data lack licenses, data cannot be considered open despite the intention of the project leaders (Bowser et al, 2020). Also, questions of legal interoperability must be highlighted if data should be merged with other datasets in the future.

Projects containing health reporting and perhaps collection of biological samples should receive special attention. For projects based outside an academic institution, it may be difficult to obtain support for an ethical review depending on the regulation and possibilities in individual countries. How participants are protected, their risk evaluated and how accidental finding disclosure will be handled are issues the project leader must consider.

Engaging specific populations in CS should be followed by clarifying their cultural needs during data collection and any resistance towards openly sharing (traditional) knowledge.

Also, it is the responsibility of the scientist to assess the consequences of data sharing and discuss this with the involved participants. Such issues may take time to investigate and should be planned - for example in a DMP or by describing a data policy.

Something to be considered early in the project is the possibility of crediting the citizen scientists for their contributed data and if certain groups of citizen scientists should be involved as co-authors on scholarly publications. As demonstrated by Hunter and Hsu (2015), applying RDA's Dynamic Data Citation Recommendation (Rauber et al, 2015) was feasible for CS project data, however, there are currently no guidelines on how to recognize citizen scientists for their contributions. A related focus area, where the library may support, is to include clearly in the descriptive metadata that data are of CS origin.

The library can build on or use the recommendations summarised above and provided in the references in Table 2 and 3. Apart from these, an international working group under the RDA has published legal interoperability recommendations that are applicable to CS projects (RDA-CODATA Legal Interoperability Interest Group, 2016). The German CS network clearly recommends communal actions to structure legal and ethical frameworks (Bonn et al, 2016) and the university libraries may be natural partners in such actions.

To summarize, the library should promote the understanding that the legal and ethical framework must be in place for data sharing and publication, and this starts with provisions for appropriate protection of privacy and sensitive information, intellectual property, relevant legislation (e.g. participant protection and laws for protection of the environment) and data rights, including licensing.

## Terms of Participation

Clear communication and alignment of expectations is a possibility for the project leader to keep the motivation and engagement of the volunteers involved in a CS project. We recommend that many of the issues addressed above be incorporated and communicated in a *Terms of Participation* directed at the volunteers. The library's role could be to support the project leader in clearly explaining the volunteers how their data are handled and used and under which conditions. It should be disclosed what are the user's rights and how personal and sensitive information is handled. Also, conditions of participant insurance could be disclosed. The information may be extracted from the project DMP, however templates for *Terms of Participation* could be developed to accommodate needs of different areas (biodiversity, health, natural science), and the policies of institutions and states.

## Incentives for continued focus on good data handling practices

RDM as a discipline develops continuously and initiatives such as the FAIR principles and the European Open Science Cloud add directions towards machine-readability and eased data access. This highlights the continuous need for quality services within RDM, but also to elucidate the cost of doing RDM – or not doing it - with the aim of securing CS data for reuse. Further, securing funding for RDM has an ethical side, since lack of funding for RDM may hamper the sustainability of a project and the possibility to maintain technologies such as platforms or apps. This may leave the efforts of the volunteers in vain and devaluate the integrity of the project.

Something lightly addressed in the included articles (August et al, 2015; Groom, Weatherdon and Geijzendorffer, 2017), but evident from the case interviews, was the incentives for not sharing data openly. Academic rewarding is generally based on the number of published scientific papers and citations; therefore, our cases are reluctant to share data before any results have been published. In contrast, volunteers may expect the project to share data openly (Crall et al, 2010) if not jeopardizing sensitive information (Ganzevoort et al, 2017). Further, several of the articles take the view of CS being a collaboration between scientists and the public and stress the importance of specifying or explaining data sharing conditions in the *Terms of Participation*. The case project leaders are very aware that the volunteers need "something in return" and different strategies have been taken from simple data download (Fyn finder marsvin) to publication of aggregated angler relevant results on website and facebook (Fangstjournalen). One solution is supporting the publication of at least metadata of the project in a repository or searchable database. This has been achieved for one of the cases since the interviews took place (Skov 2021).

Another incentive for researchers to follow good RDM practices is the possibility of having data reused and put into a new context. For example, two cases, "Fyn finder marsvin" and "Fangstjournalen" have overlapping geographical areas. The conditions of harbour porpoise and fish populations in same sea areas may generate new knowledge of ecological importance for conservation efforts. Miller-Rushing, Primack and Bonney (2012) describe how CS ecology data contribute profoundly to our understanding of the environment. However, quality contributions only emerge from efforts in securing data documentation, interoperability and access. Not securing this may have large implications for CS in terms of reputation, commitment to ethical principles or reuse (Bowser et al, 2020).

Non-scientific data quality has long been an obstacle for scientific communities and governmental bodies to embrace and reuse CS datasets (Bowser et al, 2020; Kosmala et al, 2016). The discussion on how to improve data quality is ongoing and deliberately not included in the present article. However, it is obvious that employing good RDM practices will contribute to securing contextualisation and therefore data quality. Importantly, the empowerment of collecting useful and quality data is a strong motivation factor for many volunteers (Clements et al, 2017). In the end, these could be the first points raised by the librarian when guiding upcoming CS projects.

## Library tools: the FAIR principles and the data management plan

In our literature and case study analyses, the FAIR principles acted as a framework for identifying RDM challenges (Table 1 and 5). On the other hand, the FAIR principles may be the structure to address RDM challenges of CS projects. The FAIR principles have already been explored as a central paradigm for RDM of VGI data often collected in CS projects (Bastin, Schade and Schill, 2017). The FAIR principles are adoptable by all disciplines and FAIRification of a data set can be done as a step-wise approach (Deutz et al, 2020). Our learning is that we as librarians must use the FAIR principles with a very practical approach as we have exemplified in a video directed at academic citizen scientists (Holmstrand et al, 2020). We have also summarised the findings of our article in a short guide for research librarians supporting FAIR citizen science data (Hansen, Gadegaard and Holmstrand, 2021).

The DataOne guide to writing a DMP for CS projects is another practical tool that the library may use when supporting the citizen scientist (Wiggins et al, 2013). We suggest developing DMP templates that highlights the challenges outlined above and perhaps even integrate tools and software for easing the scientist's workflow. A CS-directed DMP may act as a framework for attending relevant RDM issues and for developing the *Terms of Participation*.

# Conclusion

Many RDM challenges identified are not only specific for the CS discipline. However, particular focus should be on CS as a discipline with volunteers expecting access to - and good use of - data. These expectations may be in contradiction with current academic merits based on maximising publication numbers before sharing data. Furthermore, optimal reuse demands databases fit for containing CS provenance information and standardised data and metadata, for retrieving data subsets, and for supporting legal interoperability. Often CS projects depend strongly on data containing personal or sensitive information. Not all countries have legal, ethical or insurance policies that encompass citizen scientists in contrast to what is the case for participants in academic research projects. This should be planned and handled meticulously before launching a CS project. Last, recognising citizens for their contributions may require specific planning beforehand.

We recommend that the university library, when engaging with CS researchers, underscores the importance of clarifying legal and ethical aspects of the data collection, of developing clear *Terms of Participation* and continuously explaining the advantages of good RDM in CS projects. Many university libraries possess tools to support RDM, which can be adopted to the needs of CS projects. Given the increasing popularity of CS, the library should continuously identify or develop tools to ease the management of CS data. We conclude that advocating for writing a DMP and promoting the use of the FAIR principles, will aid CS projects throughout the data life cycle and increase the sustainability of the data.

# Acknowledgment

We are grateful to the four CS project managers for their contribution to this project. We thank Kristian Hvidtfelt Nielsen, Aarhus University, for valuable input to the manuscript.

# Funding

This article is part of a project funded by Danmarks Elektroniske Fag- og Forskningsbibliotek. The Danish RDA Node supported this article through a grant from RDA Europe 4.0 to establish national nodes and promote the work of RDA. The EU Horizon 2020 research and innovation programme funded RDA Europe 4.0 (Grant Agreement no. 777388).

# Competing Interests

The authors have no competing interests to declare.

# Tables

**Table 1. Challenges identified from literature and categorised into findability, accessibility, interoperability, reusability and research data management and infrastructures [a]**

| Reference | Aim | Findability | Accessibility | Interoperability | Reusability | Research data management and infrastructures |
|---|---|---|---|---|---|---|
| (Adriaens et al, 2015) | Apps for recording invasive species are presented and issues of data interoperability, openness and harmonisation discussed. Recommendations are provided. | Compilation of invasive species data from different regions is a growing challenge. Recommendation: "Ensure that applications generate data in a standardized format and feed into central record collection systems."<br><br>Such a system could be GBIF. Also, developing a possibility of creating alerts about new datasets to rapid-response stakeholders is encouraged. | | | Data sharing is important for managing biological invasion strategies. If shared pictures do not have a license, then linking to accompanying data is hampered. Recommendation: "Inform users about issues of intellectual property rights of records and associated media files so that this does not restrict further usage." | Apps may have overlapping functions (recording same species) which may cause confusion and competition. Long-term DM and technical updates need secure funding, also if data are used for policymaking and regulation. Recommendation: "Ensure sustainable funding or think of alternative solutions for technical updates and data verification." |
| (August et al, 2015) | Describes how new technologies are changing the study of the biological world. | UUIDs such as DOIs will secure findability. | Data curation to secure access is necessary. Moving from data sharing to data | Interoperability allows integration with other datasets and to future-proof | UUIDs ensure citeability and crediting. | Data warehouses hosting a range of different projects are a solution to secure F |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | A standard for tracing editions of a dataset should be developed. | publication with possibility to get cited may motivate to make data open access. | the data against technological changes. Use of UUIDs for taxon names. | | and A and to avoid data duplication. |
| (Bastin, Schade and Schill, 2017) | The chapter discusses how VGI data in CS and crowd-sourcing projects may be of value for individuals, institutions and decision-makers.<br>With base in the FAIR principles, VGI and generic DM principles are discussed. | Metadata for VGI are very heterogeneous, but standards do exist that can support VGI dataset to become of good quality and becoming machine-readable. Community-used terminologies require semantic mapping before they can be used across domains.<br>VGI data can only be fully appreciated if followed by a use license.<br><br>The authors describe the applicability of the FAIR principles to VGI data management. The example of GBIF is used to illustrate that cross-domain strategic thinking sustains data curation and discovery, the use of PIDs for datasets and citing, standards and taxonomies for metadata and data provenance documentation etc. | | | | Active RDM of VGI data may ensure the reproducibility necessary for data to be used for scientific and decision-making purposes.<br>Tools to document e.g. how data are packaged and what information describes accuracy is currently lacking.<br>Funding for RDM is often not considered or present in CS projects. |
| (Borda, Gray and Fu, 2020) | A scoping review on RDM practices in biomedical CS projects and an analysis of selected platforms. | Information on long-term curation and findability is not addressed publicly in scrutinized platforms. | Some, but not all, platforms state how participant data are stored and secured: e.g. genetic information is stored separately from personal and health information.<br>Some platforms may provide third parties with aggregated data. | Ensuring data quality by using standards is not addressed in scrutinized platforms. | Data processes and use of standards across data life cycle are not transparent or openly available for evaluation rendering reuse opaque, conflicted or untraceable. | |
| (Chimbari, 2017) | Lessons learned from implementing ecohealth | | Copies of data (non-electronic and | | | General recommendation: |

| | | | | | | |
|---|---|---|---|---|---|---|
| | CS projects in South Africa | | electronic) should always be transferred to PI. PI manages access rights. Community feed-back on findings is important to sustain trust and engagement.<br><br>Recommendation: A strategy including musicians or artists is recommended. | | | Develop clear RDM policies.<br><br>Tension on authorship often occurs. |
| (Clements et al, 2017) | Conclusions from a workshop on low-cost air monitoring sensors. | | | Deployment of a variety of sensors has not been followed by standardisation of data formats, units or metadata.<br>Currently, data transformation is necessary for integration. Data and metadata (e.g. time and date) format standardisation is recommended. | | There are huge prospects for saving resources and creating new knowledge by creating a large-scale data management system.<br>Currently, data are not openly shared e.g. for communities to compare.<br>The Air Sensor Workgroup works to make air sensor data FAIR: create metadata standards, software and tools in open source, and develop a data platform. |

| | | | | | | |
|---|---|---|---|---|---|---|
| (Crall et al, 2010) | A survey of CS projects working with invasive species observation is performed and obstacles for getting the most out of data are discussed. | | Access to data is hampered by concern over privacy or sensitive data (personal data, private property, and threatened, endangered species). In general, data sharing before scientific publication is wanted by survey participants. | | | Projects lack database resources and skills to share data.<br><br>Recommendation: Standardised data collection, quality assurance protocols and a national data infrastructure could improve invasive species distribution maps and detection.<br>Solutions:<br>The initiative "Global Invasive Species Information Network" aims to link online data sources.<br>Citsci.org can accommodate invasive species CS projects' data, privacy concerns and data sharing. |
| (Groom, Weatherdon and Geijzendorffer, 2017) | The paper examines openness assessed from data licensing of GBIF datasets. The relative openness of citizen science data is evaluated. | | CS data access is most often determined by CS organisations or PIs.<br>Data for GBIF are often obfuscated to accommodate privacy concerns. | | Of 1264 CS datasets only 33 has a data license. In general, usage license was more restrictive than non-CS datasets. Datasets without a license can't be used openly.<br><br>Recommendation: | 10% of dataset are from CS but constitutes 60% of all observations in GBIF. Citizen scientists may wish for recognition from community.<br>Recommendation: Recognition of contribution from |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | Data sharing may depend on funding or authorship possibilities for academic researchers. | | Organisations must implement clear licensing policies. Projects could make the volunteers choose license for their own data. | citizen scientist should be supported by data users.<br><br>Recommendations: Organisations must implement clear RDM policies.<br>Funders should recognise that quality data requires sustainability. |
| (Higgins et al, 2016) | The EU funded project, COBWEB, has researched the requirements for developing a platform for sharing environmental data from CS projects. Different solutions have been developed or suggested to accommodate the largest challenge for CS data; to make data interoperable and fit for re-use. | | The level of public access/data security is regulated i.e. to protect endangered species. User privacy is also addressed. | Existing open standards for metadata and data should be implemented. Ontologies for individual projects should match existing ontologies. | | Open source tools are developed to facilitate data collection for non-experts. The platform should offer the structure to facilitate CS data collection and improve environmental monitoring. |
| (Hunter and Hsu, 2015) | RDA's Dynamic Data Citation Working Group suggests an approach to dynamic data citation. The | CS datasets containing observations of the environment are often dynamic. | | | Volunteers are rarely cited for their contribution.<br>Recommendation: The Dynamic Citation | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | authors of this paper developed a testbed that can be used for citing sub-sets of dynamic CS datasets and also recognises the volunteers who contributed the data. | Recommendation:<br>- The underlying database must be versioned and support time stamping of changes or additions.<br>- The PID to the citable data comprises a query to the dataset and a timestamp. | | | Approach should allow contributors to the specific dataset to be recognised. | |
| (Kissling et al, 2018) | A WG addresses the need for creating a set of Essential Biodiversity Variables, when collecting biodiversity data, not only in CS projects. To enable CS data to contribute to scientific species monitoring, CS projects also needs data and workflow harmonisation. The applicability of the FAIR principles is underscored. | Datasets must be findable and citable. | Data access restrictions may severely hamper quality control, data aggregation and reuse. | CS data needs rich metadata to assure quality and reuse. Data must be machine-readable. | Documentation and licensing information must accompany published data. Legal interoperability is required for automated workflows and is necessary for data aggregation which is widely used in biodiversity monitoring. However, different licenses for different datasets may restrict use of aggregated datasets, therefore, CC0 and CC BY are endorsed. | |
| (Owen and Parker, 2018) | The authors describe how CS data can be used for EPAs and other policy-making bodies. | | | Metadata are necessary for data quality and for use by EPAs. | EPAs can use CS data of certain quality. | Authors encourage EPAs to offer infrastructure to CS projects. |

| | | | | | | |
|---|---|---|---|---|---|---|
| (De Pourcq and Ceccaroni, 2018) | The blogpost describes the advantages of and organisations behind creating a data and metadata standard for CS projects. | | | Incompatible data handling hampers data reuse. Reuse of project structures and methods overall is also unlikely if not transparent and following minimum standards.<br><br>The International Data and Metadata Working Group and the CS COST Action will launch a standard on key elements and concepts of CS projects. Guidelines for its implementation will be provided. | | Authors encourage good RDM practices in CS projects to facilitate better data quality. |
| (Pulsifer, Huntington and Pecl, 2014) | This editorial introduces a special issue of *Polar Geography* on the challenges and prospects for better inclusion of local and indigenous observations in environmental knowledge. | | Observations may contain sensitive information about a people or region that they may not want to share openly. | | | Access to RDM systems in remote communities may be difficult. But they can link observations from different stakeholders. RDM is not only a question of technical and methodological aspects, but must encompass local culture and economy. |

| | | | | | | |
|---|---|---|---|---|---|---|
| (Runnel and Wijers, 2019) | The WG report addresses issues about managing natural history collections data used in CS projects. | CS portals rarely allow searching for collection-based projects. Metadata standards should facilitate this. | | Metadata standards should be adapted to contain information describing natural history collection data. CS project metadata should reveal if data originated from a collection. This will aid transparency for policy makers and recognition of participants. | | |
| (Schade, Tsinaraki and Roglia, 2017; Schade and Tsinaraki, 2016) | Survey report and related publication on data management in CS projects. | | Observation: Interest to share data is large, but several projects do not provide immediate access.<br>Many projects cannot guarantee sustained or any access to data. Funding for this may be insufficient. | Observation: Data and metadata standards are not applied in many projects. Funding for managing this may be insufficient. | Observation: Licensing is often determined only late in projects and may cause confusion. | Identified DM needs: Promotion of Open Data, Open Science, data preservation, existing infrastructures, development of standards through guidelines and best practices in relevant communities. |
| (Sheppard, Wiggins and Terveen, 2014) | Proposes a model for data provenance/workflow in field sampling and processing. | | | | To make data reusable, documentation and metadata are necessary to track changes to data (provenance), e.g. cleaning, re-entry, new/changed protocol | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | for task definition/sampling. | |
| (Simonis, 2018) | Proposes a standard model for describing CS data, so they become interoperable and reusable. | | | The model builds on existing standards. Model is based on resolvable URLs for semantics/identifier to make raw data meaningful for all and machine-readable. | | |
| (Williams et al, 2018)<br><br>- Refer to Table 2 for more data from this reference. | The chapter addresses which factors should be considered to maximize the use and impact of CS data. | | Data accessibility should be considered early in project. | Few CS projects adopt standards for web services or data encodings, because the benefits of sharing data is unclear or because resources to do it are lacking.<br><br>Interoperability is not only important for machine-interaction, but also for human-machine and community interactions.<br><br>Specific metadata standards can be useful for different organisation, e.g. DCAT for open governmental data. | Preparing CS data for reuse secures the long-term value, therefore consider<br>- which contributions are subject to IPR<br>- data ownership<br>- data use license<br><br>Contextualising data with metadata, including descriptions of their purpose and methods of creation, allows users to evaluate the reuse and possibility to integrate with other datasets.<br><br>Data provenance/processing can be difficult to document and therefore | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | Semantic interoperability represents the highest level of interoperability for data exchange, quality and sharing. | understand for other users. | |

[a] Abbreviations: CS, citizen science; DCAT, Data Catalogue Vocabulary; DMP, data management plan; DOI, digital object identifier; EPA, environmental protection agency; GBIF, Global Biodiversity Information Facility, PID, persistent identifier; PI, principal investigator; RDA, Research Data Alliance; RDM, research data management; UUID, universally unique identifier; VGI, Volunteered Geographic Information; WG, working group.

**Table 2. Ethical and legal challenges identified in literature [a]**

| Reference | Aim | Content summary |
|---|---|---|
| (Anhalt-Depies et al, 2019) | A framework is conceptualised in which tension in CS is discussed. Privacy policies of 20 projects are reviewed and recommendations offered. | CS data may contain private or sensitive information, e.g. landownership, personal information or pictures of persons, location of endangered species.<br>Privacy-related policies were very different in content and not always project-specific.<br>Recommendations:<br>- During project development, identify potential tensions between data quality, privacy protection, resource security, transparency, and trust in consultation with stakeholders.<br>- Develop a privacy policy or volunteer agreement that addresses these tensions and is consistent with existing guidelines<br>- Develop a data sharing policy that clearly states any restriction on data sharing; consider impacts on resource security and volunteer privacy in determining restrictions, and plan for what to do if a difficult scenario should arise (i.e. detection of illegal activity)<br>- Practice iterative evaluation of policies and practices in use to assess their impact on the ability to achieve program goals<br>- Develop a process for soliciting regular feedback from participants |
| (Bowser et al, 2014) | Through examples, the article addresses legal and policy considerations that protect participant privacy in CS. US | Five recommendations are provided:<br>- Determine which data points you can and cannot compromise on in terms of precision, public visibility, and data sharing; clearly state these decisions, and implement the supporting technologies (fuzzing locations, anonymizing identities, etc.).<br>- Give ample notice of privacy choices. Explain the circumstances under which normal participation could be a risk to personal privacy. Inform volunteers who will review their data for quality control. |

| | | |
|---|---|---|
| | law and policy is primary offset for article. | - Give volunteers the option to hide certain data points and locations from public view, or have data publicly visible but attributed anonymously.<br>- Allow volunteers to delete and modify their data—both traditional personal information and submitted data that may contain information “about” the volunteer.<br>- Require only minimum personal data about volunteers. Demonstrate the value of the data you collect, and explain who will be able to see it. Multilevel access control that considers different stakeholders’ roles and needs may be appropriate. |
| (Bowser et al, 2017) | A qualitative study of the privacy concerns of CS study managers and volunteers.<br>It is suggested how to design data and information flow and design supporting technologies in CS projects. | Participants evaluate privacy risk in the context of the project. They focus on openness and sharing for personal and collective benefits.<br>Current research regulations may not sustain the culture in CS projects, where concern for privacy is sometimes outweighed by incentives for data sharing.<br><br>Recommendations:<br>- Minimise personal data collection to sustain trust of volunteers.<br>- Support privacy through design: build-in notifications, filter data upon submission.<br>- Teach volunteers about the data flow. |
| (Ganzevoort et al, 2017) | A questionnaire survey of CS biodiversity volunteers’ motivation for collecting data and their views on data sharing and ownership. | Half the respondents view data as a public good, but only few support unconditional sharing. Data should be used for nature protection and with great respect.<br>69% would like insight to the use of their data.<br>Ca. 40% would like to be cited by name when their data were used. |
| (Guerrini et al, 2018) | The article discusses issues around intellectual property rights, research integrity and participant | <u>Intellectual property:</u><br>Volunteers retain the IPR to any copyrightable work they produce. Recommendation: Use CC licenses and make copyright agreements in the projects.<br>Patent assignment as known from employer-employee discoveries rarely occurs in CS. Thus, CS inventors can exclude projects in using the CS invention. Disagreement on license or patent may occur.<br>An obstacle is that CS organisations often don’t have funding to negotiate IPR control. |

| | | |
|---|---|---|
| | protection in CS projects. These issues are not always or not clearly regulated by laws or institutional policies. | One-way material transfer agreements could be adapted to promote CS sharing, but may be complex to handle. Transparency and clear IPR terms is recommended in CS collaborations.<br><br>Recommendation: Contracts with volunteers can be made that render project leaders the patent rights or that share the patent right between project leader and CS inventor(s).<br><br><u>Research integrity:</u><br>May be challenged in CS projects if e.g. purpose is biased towards promoting or preventing a community intervention. US federal sponsored CS data must be made openly available to increase transparency. Such laws are not widespread in other countries. Research integrity often relies on peer-reviewing when publishing articles.<br>CS volunteers cannot disclose conflict of interests.<br><br>Recommendation: Making protocols and data openly available promotes research integrity.  Giving volunteers the possibility to stay anonymous is more important than their disclosure of conflicts of interest.<br><br><u>Participant protection:</u><br>Volunteers are not protected by laws normally regulating research subjects. Projects may not be reviewed by institutional boards if founded outside academia. Participant risks may not be disclosed in terms of participation.<br><br>Recommendations: Community advisory committees may review studies. If funding is available for projects outside academia, IRB evaluation could be obtained. Further efforts are necessary to evaluate if laws can be extended to CS or if specific policies should be created together with citizen scientists. |
| (Oberle et al, 2019) | From the example of a Canadian CS project, ethical review of CS projects is discussed | The responsibilities of the IRB review is to protect subject from harm, but generally citizen scientists are “research assistants” rather than “research subjects” and do not fall under IRB reviews.<br>It is suggested that CS projects are reviewed by the legal or public relations department rather than the IRB. However, an initial evaluation of harm from an ethical perspective before deciding for an IRB review could also be a solution. |
| (Patrick-Lake and Goldsack, 2019; Wiggins and Wilbanks, 2019) | A connected editorial and article. The complexity of issues that CS projects in health | The definition of what CS encompasses is often blurred. The current technology facilitates new possibilities of data collection, which is “CS-like”. Thus, in several projects, participants act more as research subjects than active citizen scientists.<br>Concerns about participant ethics and protection is valid, because the risks to participants delivering health data is not necessarily addressed. |

| | | |
|---|---|---|
| | and biomedical need to consider are discussed and concerns exposed. | Projects focussing on intervention rather than observation may raise more ethical issues and pose larger risks for participants.<br>CS projects originating from outside academic institutions do not always follow academic regulations and policies.<br>Informed consent can be obscured for participants engaging in data collection that is CS-like.<br>Non-researchers may initiate research where data are delivered to third-parties.<br>Direct publication of non-academic CS data without peer-review and quality control can lead to misinformation.<br>Current ethical frameworks are aimed at handling evaluating risks and protecting participants, and not fit for helping autonomous and engaged co-researchers (citizens). |
| (Resnik, Elliot and Miller, 2015) | The authors discusses the ethical challenges occurring in CS as a collaboration between laypeople and scientists. | <u>Research integrity</u><br>Research integrity could be compromised in CS projects, where data collectors or project initiators are aiming to address a community-issue of particular concern. Projects may also be funded by organisations or corporate funds with e.g. lobbying, legal or political interests. Both financial and non-financial conflicts of interest should be addressed in the project, both in the beginning and when publishing data and results. Disclosure of conflict of interest could be performed individually or as a group.<br><u>Access:</u><br>Data sharing will allow others to evaluate data independently. Potential policies for CS projects on conflicts of interest should, however, not prevent communities for engaging in research that may help them fight e.g. environmental injustice.<br><br>Data sharing allows others to reuse, discuss and give feedback. Data must be de-identified if containing information on human research subjects. Citizens should be clearly informed of the expected sharing of data (who, when, why).<br>Data ownership and IPR issues may arise if communities expect to have some control over the gathered data. Agreements should be clear and updated regularly with the volunteers. Sharing of culturally-embedded knowledge should be handled with respect.<br><br>Exploitation of volunteers could occur if the volunteers do not receive a share of benefits potentially obtained by the research they participated in. The scientist should aim at sharing IPR, authorship, formal recognition, education or monetary value.<br>Safety of volunteers should be considered.<br><br>Co-authorship should be considered for volunteers providing substantial contributions to the study, but may often fall outside the recommendations of ICMJE. The authors encourage credit in the acknowledgment section and sharing of results.<br>The concept of CS may be used misleadingly, e.g. volunteers may serve more as data collectors or research subjects than active participants. |

| | | |
|---|---|---|
| (Riesch and Potter, 2014) | Qualitative study of CS researchers on methodological, episthemiological and ethical issues. | There is consensus that a CS project should at least be transparent with the data it collects, what it is being used for, and how to keep citizens updated on the process.<br><br>The question on how citizens should be credited is raised. Data are produced by the public, so ownership is a question to consider. |
| (Rothstein, Wilbanks and Brothers, 2015) | The article discusses how newly emerging, technology-enabled, unregulated CS health research poses a substantial challenge for traditional research ethics. In the US, CS projects set up by private persons are not regulated as is company- and academic-driven research. | A: There are no data sharing or publication obligations for private CS projects.<br><br>IR: Without review, the validity of data and results may not be scrutinized or assessed.<br><br>Projects may not have institutional review, and ethical approval, which can oversee recruitment procedures, participant eligibility and informed consent. Requirements for protection of privacy and confidentiality remain unclear.<br>How can child participants be monitored by legal guardians?<br>Should incidental findings be disclosed and how? |
| (Tauginienė, 2019) | The article aims to address ethical aspects of CS projects with focus on research integrity. | No consensus on CS authorship or attributions exists.<br><br>To increase transparency, informed consent should address the relationship between scientist and citizen and the citizen's role in the research. The scientist must act socially responsibly by informing society of methods, tools, data and knowledge. |
| (Ward-Fear et al, 2020) | The article discusses if and how citizen scientists should be | Current scientific authorship criteria excludes citizens to be attributed co-authorship.<br>The authors propose implementation of group co-authorship to cohorts of non-professional scientists. |

| | | |
|---|---|---|
| | included as co-authors. | |
| (Williams et al, 2018)<br><br>- Refer to Table 1 for more data from this reference. | The chapter addresses which factors should be considered to maximize the use and impact of CS data. | Primary IPR considerations for CS: (1) “background IPR” – How will knowledge and data be used and under what restrictions; and (2) “foreground IPR” –how will the project allow access to the knowledge and data.<br><br>Personal privacy must be protected, i.e. personal information and location details.<br>Protection of security for objects collected must be considered, e.g. endangered species or unintentional photo capture of persons or secondary objects.<br><br>Handling of IPR and privacy should be described in *Terms of participation.* |

[a] Abbreviations: CS, citizen science; CC,creative commons; , IPR, intellectual property rights;  IRB, institutional review board; ICMJE, the International Committee of Medical Journal Editors

## Table 3. Identified tools, roadmaps and guidelines for research data management of citizen science[a]

| **Reference** | **Aim** | **RDM content in reference** |
|---|---|---|
| (US GSA) | A short toolkit from the U.S. federal government on managing CS data | |
| (Bonn et al, 2016) | A Green Paper presenting the understanding, requirements and potential of CS in Germany and is a roadmap towards 2020. Guiding principles are also presented. Two chapters discuss data management of and the legal and ethical framework for CS.<br>The recommendations for action are listed here: | General RDM:<br>• Establish framework conditions for securing data quality<br>• (Further) develop automated data validation and statistical methods to analyse Citizen Science data<br>• Establish framework conditions for adaptive data management:<br>• Enable an open-science policy (open access and open source) for Citizen Science data<br>• Establish and implement the use of a standardised citation format for Citizen Science data<br>• Establish and implement guidelines for quotable metadata<br>• Develop guidelines for harmonising different data sources without loss of information content or data source traceability<br>• Develop long-term repositories for Citizen Science project data<br>• Provide support for such repositories in the long term<br>• Integrate and support established structures for implementing data management, e.g. in scientific archives, libraries and collections<br>• Develop a legal framework for handling intellectual property rights to enable the recognition of new inventions as communal goods<br>• Establish coordination and data information offices to assist with data issues when designing and analysing Citizen Science project results. |

| | | |
|---|---|---|
| | | Ethical and legal:<br>• Develop proposals for dealing with intellectual property rights, data protection and monitoring of compliance with regulations<br>• Draft action guidelines on the topics "data openness", "intellectual property" and "data protection" for Citizen Science project initiators and participants<br>• Develop standards for collaboration agreements between institutionally affiliated and independent Citizen Science partners<br>• Set up extended insurance coverage for volunteers actively participating in Citizen Science programmes<br>• Clarify and review ethical issues relating to all aspects of Citizen Science |
| (Disney et al, 2017) | Presentation of the CS project tool, anecdata.org – an online platform for CS project to collect, manage and share environmental data. | Works as a repository to share and download data openly.<br><br>May be connected to SciStarter.com in the future. Apparently does not support other RDM functions than data storage and sharing. |
| (Forest Service, 2018) | A guide from US Forest Service for CS projects in order to make data of good quality available to the agency. Chapter 4 mentions DM shortly. | Data should be made available to Forest Service staff. |
| (Greshake Tzovaras et al, 2019) | A new platform, Open Humans, is presented. The platform is open for personalised data collection (e.g. health data), but allows participants to control sharing. The platform can be used for CS and academic research. | The article present challenges for participatory science within humanities, sociology and medicine:<br>- Accessing data in commercial environments (e.g. apps)<br>- Health data are stored in "silos", e.g. managed by national institutions<br>- Ethical concerns over use of personal data<br><br>Participants can upload data collected elsewhere and manage which projects on Open Humans that can access the data.<br>Data can be re-used in as much as possible under the control of the participant.<br><br>Members share notebooks (code for data analyses) that allows analysing the individuals own data, i.e. notebooks are interoperable and reusable<br>The open source for the platform has allowed communities to write own expansions and data importers. |
| (Heigl et al, 2018) | The CS Network Austria has defined a set of quality criteria for projects | FAIR: |

| | | |
|---|---|---|
| | wishing to be listed on the Austrian CS platform, *Österreich forscht*. The criteria are also formulated as questions, which project leaders must answer. Platform coordinators and a WG read the answers and provide feedback and support if deemed necessary.<br>Criteria relevant for RDM are listed here. | - All data and metadata is made publicly available, provided there are no legal or ethical arguments against doing so.<br>- The results are published in an open-access format, provided there are no legal or ethical arguments against doing so.<br>- The results are findable, reusable, comprehensible and transparent.<br><br>RDM:<br>- Prior to data collection, all projects must have established a data management plan which conforms to the European General Data Protection Regulation<br><br>Ethical and legal issues:<br>- The project must follow transparent ethical principles in compliance with ethical standards, such as obtaining informed consent from participants or the parents of participating children, among others.<br>- Clear information on data policy and governance (regarding personal and research data) must be published within the project, and participants must consent to this information prior to participation. |
| (Parthenos) | An online course/resource for CS in (digital) arts and humanities. One module focuses on DM planning of CS or crowd-sourcing projects. Additional modules deals with research infrastructures and ethics | Recommendations:<br>- Know what you data will be, and how you will use it, to ensure you are compliant with GDPR and ethical standards<br>- Use appropriate standards to model your data<br>- Use a data management plan to help structure your thinking |
| (Pettibone et al, 2016) | A guide for practitioners on citizen science as practised in Germany. One chapter is on data and legal considerations. | Data should be secured for long-term use in permanent infrastructure<br>Data rights must be determined.<br>Reusability must be ensured through clarity of data and use of appropriate metadata.<br><br>DM must be transparent and comply with legal requirements.<br><br>Ethical and legal issues:<br>The legal framework must be in place, considering copyright, data rights, privacy, personal data and relevant legislation (e.g. laws for protection of the environment) |
| (Sturm et al, 2018) | Recommendations from workshops on principles for mobile apps and platforms in CS projects. It is | The workshop identified and provided recommendations for RDM challenges related to securing interoperability and data management: |

| | | |
|---|---|---|
| | acknowledged that the recommendations can be used for CS projects in general. | Index apps and platforms to facilitate reuse.<br>Data sharing and use of open source for code base is encouraged. Consider data privacy.<br>Use standards for software design and for data and metadata. Use UUID for all observations and data points.<br>For reuse of apps and platforms, include metadata for license, documentation and modifications. Provide technical support for the app/platform.<br><br>Recommendations on securing sustainability of the project, data protection, participant privacy and IPR (incl. national/regional differences) are also provided. |
| (Tweddle et al, 2012) | A guide to CS written on behalf of the UKEOF, i.e. directed at environmental sciences. A few advices on RDM is included. | Store data in well-known repositories. Make data available electronically. Data sharing with relevant organisations is encourage, since they often can provide data storage.<br><br>Ethical and legal issues:<br>IPR and data protection requirements must be considered. |
| (UKEOF's Advisory Group, 2013) | A pamphlet that shortly explain seven principles to ensure quality data and good data management of CS projects. | Consider the data requirements<br>Manage volunteers to get the best data<br>Ensure data quality<br>Harness new technologies<br>Manage data effectively<br>Report and share data<br>Evaluate to maximise data value |
| (US EPA, 2019) | Handbook by US EPA that addresses how to ensure quality, documentation and data management of CS projects. | The handbook contains detailed<br>- advices and templates for documentation and data reuse<br>- advices and a template for writing a DMP |
| (Wang et al, 2015) | Presentation of the CS project tool, CitSci.org | CitSci.org is a customizable platform that allows users to collect and generate diverse datasets.<br>It contains standardised metadata necessary for data exchange and quality assurance.<br>A web-based DM feature is included in tool.<br>The tool includes documentation of permissions, privacy and security of information. |
| (Wiggins et al, 2013) | DataOne WG report on introduction to data management of CS projects. The report function as a tool for RDM. | The document<br>- introduces the data life cycle<br>- provides best practices and recommendations in each step of this life cycle<br>- identify key opportunities and challenges in DM |

| | | |
|---|---|---|
| (Wolf et al, 2019) | ONC is university-based and operates ocean observatories and repositories services. ONC has developed a DM system and the article presents how ONCs best practices and services for DM is applied to a CS project in the entire data life cycle, rendering CS data FAIR. Can be used as a tool/guideline for RDM. | The document describes how ONC implements best data management practices throughout the data life cycle. Can be used as a tool/guideline for RDM. |

[a] Abbreviations. CS, citizen science; DM, data management; DMP, data management plan; RDM, research data management; IPR, intellectual property rights; OCN, Ocean Networks Canada; UKEOF; the UK Environmental Observation Framework; US EPA, United States Environmental Protection Agency; WG, working group

**Table 4. Information about projects in case study**

| **Project title** (Translated) | **Homepage and start year** | **Purpose** | **Citizen scientist** | | | **Researchers** | | **Dissemination to the public** |
|---|---|---|---|---|---|---|---|---|
| | | | Prerequisites | Involvement | Outcome | Benefits from using citizen science method | Outcome | |
| **Fyn finder marsvin** (Funen finds harbour porpoises) | https://www.sdu.dk/da/forskning/forskningsformidling/citizenscience/fyn+finder+marsvin<br><br>2019 | Distribution of harbour porpoises in the inner Danish waters: Spatial, seasonal, and females with young cubs. | All persons with a cell phone. | Observations collected via mobile app. | The participant will get an understanding of how many resources population registration requires by conventional scientific method. Learn about harbour porpoise biology. | Large spatial coverage and large data volume | Publicity in the media.<br>Research data, merit, and a basis for management and conservation | Website with observations data on university and partner website.<br>Radio interviews and articles in popular science magazines. |
| **Livet med demens** (Life with dementia) | https://www.sdu.dk/da/forskning/forskningsformidling/citizenscience/lidem<br><br>2019 | The purpose is to create a centre for dementia, under which research projects can be developed and run in collaboration with citizens, professionals, municipalities and scientists. | Patients with dementia, their relatives, caretakers and other professionals can participate. | The participants' knowledge on how to live a life with dementia will be actively used. | Larger inclusion of relatives and caretakers.<br>Increased quality of life for relatives and patients. Better treatment of patients. | More knowledge about what works best, to increase the quality of life for both patients and relatives.<br>To put dementia on the political agenda. | New methods will be tested and documented in order to create better treatment and increase the quality of life. | Physically by small theatre productions, material for website and directly to participating municipalities.<br>Scholarly publication and conferences. |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Fangstjournalen** (CatchLog) | https://fangstjournalen.dtu.dk/<br><br>2016 | Better knowledge on fish populations in Danish waters. | All persons with cell phone and/or web access with an interest in fish and aquatic environment. | Collect information about fish from fishing trips via app or browser. Collect observations e.g. about large mammals from aquatic environment. | Logbook of own fishing trips, possibility to show catches to others. The app gives information about current location fishing restrictions. | Data could not be obtained by other methods and provide large spatial coverage and data volume. | Research data, merit, and a basis for management and conservation | Continuous publication of news and data on website and facebook. Scholarly publications and conferences. |
| **Masseeksperiment 2019** (Mass Experiment 2019) | https://naturvidenskabsfestival.dk/tildinundervisning/masseeksperiment-2019-plastforurening-i-vand<br><br>2019 | Distribution of plastic litter in the Danish terrestrial environment. | School and high school children (grades 0-9 and 10-12 in DK). | Collect, classify, and count plastic litter | Can be part of school teaching curriculum: Insight into the problem of plastic pollution in the Danish environment. | Large spatial coverage and large data volume. | Research data and merit. | Report is published and a scholarly paper is submitted. |

**Table 5. Solutions and challenges with research data management and infrastructures, FAIR and ethical and legal issues**

Data is extracted from interviews with the principal investigator of projects in case study[a].

| **Project title** (Translated) | **Research data management and infrastructures** | **Findability** | **Accessibility** | **Interoperability** | **Reusability** | **Ethical/Legal issues** |
|---|---|---|---|---|---|---|
| **Fyn finder marsvin** (Funen finds harbour porpoises) | There was no initial intention to write a DMP, though the university's Open | Results can be found through the project homepage, and in an open repository[b].<br>A DOI and simple | All sightings available through website. The full data set is uploaded to Zenodo at intervals. | Data and metadata are not defined by ontologies. Data consist of the porpoise sightings (date, | Data are published in Zenodo under the CC BY 1.0 license, but are not accompanied | Only locations for porpoise sightings are shared, data do not contain any personal information. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Science Policy mandates one.<br>PI not aware of the FAIR principles. | administrative metadata are assigned to the data in the repository. | | number and location), are of very simple structure and can be downloaded in csv format. | by provenance documentation. | |
| **Livet med demens** (Life with dementia) | DMP may be written for individual projects. The centre is currently developing activities.<br>PI not aware of the FAIR principles. | Some data could be made available, but of course not patient data. | | | | Patient level data are highly sensitive.<br>Mapping data showing how municipalities are working with patients can be shared. There are also qualitative “data” that could be shared with consent. |
| **Fangstjournalen** (CatchLog) | To write a formal DMP was not a recommendation at the time of project start. A DMP would have been useful.<br><br>Data structure not initially designed for a repository.<br><br>PI not aware of the FAIR principles and the institutional data repository. | Aggregated results can be found through the app and project homepage, but data not available in an open repository.<br><br>Currently no PID or administrative metadata are assigned to the data.<br><br>(A metadata record is available in an open repository since 2021[c].) | Data are stored in local database.<br>Datasets can be shared as a copy after cleaning for personal data - no direct access to data. | Some standards are used for structural metadata and data formats.<br><br>Machine readable identifiers are not assigned to data.<br><br>PI has suggested a standard for angler projects[d]. | PI sees great potential with merging data from other aquatic and environmental sources.<br><br>Data quality is high and documented, but not publicly available yet.<br>Manual work needed for data cleaning and assigning metadata before any kind of sharing.<br><br>PI interested in sharing and licensing | GDPR is a major issue - as the ‘fear’ of breaking GDPR rules hinders the willingness/courage to share data.<br><br>Processes for anonymising data before publication/sharing needs to be defined and cleared. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | data through the institutional repository, but with embargo until results have been published in scientific articles. | |
| **Masseeksperiment 2019** (Mass Experiment 2019) | To write a DMP was not a recommendation at project start, but would have been useful. Data structure not initially designed for a repository. Raw data stored at Astra (the national Centre for Learning in Science, Technology and Health in Denmark). PI not aware of the FAIR principles. | When an article presenting the results was submitted, data were uploaded to Zenodo and DOI and metadata were added[e]. | Data published in Zenodo[c], however with personal data removed (GPS coordinates, school names etc.). | Currently no known standards for this type of data (format, metadata) except that plastics were classified according to [f]. | When data is published in an open repository, the datasets will be kept as original as possible but with anonymization. The data are published as an Excel file with no provenance information under the CC BY 4.0 license. | No personal data involved. School class data and spatial data (GPS coordinates) are removed. |

[a] Abbreviations: DMP, data management plan; DOI, digital object identifier; PI, principal investigator; PID, persistent identifier.
[b] (Wahlberg, 2020)
[c] (Skov, 2021)
[d] (Venturelli, Hyder and Skov, 2017)
[e] (Syberg, 2020)
[f] Annex 1 in (Hanke et al, 2020)

# References

**Adriaens**, **T**, **Sutton-Croft**, **M**, **Owen**, **K**, **Brosens**, **D**, **Van Valkenburg**, **J**, **Kilbey**, **D**, **Groom**, **Q**, **Ehmig**, **C**, **Thürkow**, **F**, **Van Hende**, **P**, and **Schneider**, **K.** 2015. Trying to engage the crowd in recording invasive alien species in Europe: experiences from two smartphone applications in northwest Europe. *Management of Biological Invasions*, 6(2): 215–225. DOI: https://doi.org/10.3391/mbi.2015.6.2.12

**Anhalt-Depies**, **C**, **Stenglein**, **J L**, **Zuckerberg**, **B**, **Townsend**, **P A**, and **Rissman**, **A R.** 2019. Tradeoffs and tools for data quality, privacy, transparency, and trust in citizen science. *Biological Conservation*, 238: 108195. DOI: https://doi.org/10.1016/j.biocon.2019.108195

**August**, **T**, **Harvey**, **M**, **Lightfoot**, **P**, **Kilbey**, **D**, **Papadopoulos**, **T**, and **Jepson**, **P.** 2015. Emerging technologies for biological recording. *Biological Journal of the Linnean Society*, 115(3): 731–749. DOI: https://doi.org/10.1111/bij.12534

**Bastin**, **L**, **Schade**, **S**, and **Schill**, **C** 2017 Data and Metadata Management for Better VGI Reusability. In: Foody, G, Fritz, S, Mooney, P, Olteanu-Raimond, A-M, Fonte, C C, and Antoniou, V (eds.), *Mapping and the Citizen Sensor*. London: Ubiquity Press. pp. 249–272. DOI: https://doi.org/10.5334/bbf.k

**Bielefeld University Library** n.d. What is BASE? *https://de.base-search.net/about/en/index.php*. Available at https://de.base-search.net/about/en/index.php [Last accessed 14 September 2020].

**Bonn**, **A**, **Richter**, **A**, **Vohland**, **K**, **Pettibone**, **L**, **Brandt**, **M**, **Feldmann**, **R**, **Goebel**, **C**, **Grefe**, **C**, **Hecker**, **S**, **Hennen**, **L**, **Hofer**, **H**, **Kiefer**, **S**, **Klotz**, **S**, **Kluttig**, **T**, **Krause**, **J**, **Küsel**, **K**, **Liedtke**, **C**, **Mahla**, **A**, **Neumeier**, **V**, **Premke-Kraus**, **M**, **Rillig**, **M C**, **Röller**, **O**, **Schäffer**, **L**, **Schmalzbauer**, **B**, **Schneidewind**, **U**, **Schumann**, **A**, **Settele**, **J**, **Tochtermann**, **K**, **Tockner**, **K**, **Vogel**, **J**, **Volkmann**, **W**, **von Unger**, **H**, **Walter**, **D**, **Weisskopf**, **M**, **Wirth**, **C**, **Witt**, **T**, **D**, **W**, and **Ziegler**, **D** 2016 *Green Paper Citizen Science Strategy 2020 for Germany.* Available at https://www.buergerschaffenwissen.de/sites/default/files/assets/dokumente/gewiss_cs_strategy_englisch.pdf

**Borda**, **A**, **Gray**, **K**, and **Fu**, **Y** 2020 Research data management in health and biomedical citizen science: practices and prospects. *JAMIA Open*, 3(1): 113–125. DOI: https://doi.org/10.1093/jamiaopen/ooz052

**Bowser**, **A**, **Cooper**, **C**, **De Sherbinin**, **A**, **Wiggins**, **A**, **Brenton**, **P**, **Chuang**, **T-R**, **Faustman**, **E**, **Haklay**, **M (Muki)**, and **Meloche**, **M** 2020 Still in Need of Norms: The State of the Data in Citizen Science. *Citizen Science: Theory and Practice*, 5(1): 18. DOI: https://doi.org/10.5334/cstp.303

**Bowser**, **A**, **Shilton**, **K**, **Preece**, **J**, and **Warrick**, **E** 2017 Accounting for Privacy in Citizen Science. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. New York, NY, USA: ACM. pp. 2124–2136. DOI: https://doi.org/10.1145/2998181.2998305

**Bowser**, **A**, **Wiggins**, **A**, **Shanley**, **L**, **Preece**, **J**, and **Henderson**, **S** 2014 Sharing data while protecting privacy in citizen science. *Interactions*, 21(1): 70–73. DOI: https://doi.org/10.1145/2540032

**Bürger schaffen Wissen** n.d. Available at https://www.buergerschaffenwissen.de/ [Last accessed 28 June 2020].

**Chimbari**, **M J** 2017 Lessons from implementation of ecohealth projects in Southern Africa: A principal investigator's perspective. *Acta Tropica*, 175: 9–19. DOI: https://doi.org/10.1016/j.actatropica.2016.09.028

**Clements**, **A L**, **Griswold**, **W G**, **RS**, **A**, **Johnston**, **J E**, **Herting**, **M M**, **Thorson**, **J**, **Collier-Oxandale**, **A**, and **Hannigan**, **M** 2017 Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary). *Sensors*, 17(11): 2478. DOI: https://doi.org/10.3390/s17112478

**Corrall**, **S**, **Kennan**, **M A**, and **Afzal**, **W** 2013 Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research. *Library Trends*, 61(3): 636–674. DOI: https://doi.org/10.1353/lib.2013.0005

**COST Action CA 15212** 2019 Workshop Report WG5: On citizen-science ontology, standards and data. Available at https://cs-eu.net/news/workshop-report-wg5-citizen-science-ontology-standards-and-data [Last accessed 5 June 2020].

**Cox**, **A M** and **Corrall**, **S** 2013 Evolving academic library specialties. *Journal of the American Society for Information Science and Technology*. DOI: https://doi.org/10.1002/asi.22847

**Crall**, **A W**, **Newman**, **G J**, **Jarnevich**, **C S**, **Stohlgren**, **T J**, **Waller**, **D M**, and **Graham**, **J** 2010 Improving and integrating data on invasive species collected by citizen scientists. *Biological Invasions*, 12(10): 3419–3428. DOI: https://doi.org/10.1007/s10530-010-9740-9

**CS WG OSPP** 2018 Recommendations of the OSPP on Citizen Science. Available at https://ec.europa.eu/research/openscience/pdf/citizen_science_recomendations.pdf

**CSNA** n.d. Österreich forscht. Available at https://www.citizen-science.at [Last accessed 1 September 2020].

**Danish Parliament** 2011 *Lov om videnskabsetisk behandling af sundhedsvidenskabelige forskningsprojekter*. Denmark: retsinformation.dk. Available at https://www.retsinformation.dk/eli/lta/2011/593

**De Pourcq**, **K** and **Ceccaroni**, **L** 2018 On the importance of data standards in citizen science | COST Action CA15212. *Citizen Science Cost Actio*. Available at https://cs-eu.net/blog/importance-data-standards-citizen-science [Last accessed 12 May 2020].

**Deutz**, **D B**, **Buss**, **M C H**, **Hansen**, **J S**, **Hansen**, **K K**, **Kjelmann**, **K G**, **Larsen**, **A V**, **Vlachos**, **E**, and **Holmstrand**, **K F** 2020 How to FAIR: a website to guide researchers on making research data more FAIR. *Zenodo*. DOI: https://doi.org/10.5281/ZENODO.3712065

**Disney**, **J**, **Bailey**, **D**, **Farrell**, **A**, and **Taylor**, **A** 2017 Next Generation Citizen Science Using Anecdata.org. *Maine Policy Review*, 26(2): 70–79. Available at https://digitalcommons.library.umaine.edu/mpr/vol26/iss2/15

**ECSA** 2015 Ten principles of citizen science. London. Available at https://ecsa.citizen-science.net/sites/default/files/ecsa_ten_principles_of_citizen_science.pdf [Last accessed 14 July 2019].

**European Commission** n.d. Open access & Data management. *H2020 online manual*. Available at https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm [Last accessed 1 September 2020].

**Forest Service** 2018 Design Your Project & Data Management Plan. In: Tamez, M, Merriman, D, and Zimmerman, N (eds.), *Forest Service Citizen Science Project Planning Guide*. FS-1.0. Forest Service, United States Department of Agriculture. Available at https://www.fs.usda.gov/working-with-us/citizen-science/ch-4-design-your-project-data-management-plan

**Ganzevoort**, **W**, **van den Born**, **R J G**, **Halffman**, **W**, and **Turnhout**, **S** 2017 Sharing biodiversity data: citizen scientists' concerns and motivations. *Biodiversity and Conservation*, 26(12): 2821–2837. DOI: https://doi.org/10.1007/s10531-017-1391-z

**GO FAIR** n.d. F2: Data are described with rich metadata. Available at https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/ [Last accessed 14 September 2020].

**Greshake Tzovaras**, **B**, **Angrist**, **M**, **Arvai**, **K**, **Dulaney**, **M**, **Estrada-Galiñanes**, **V**, **Gunderson**, **B**, **Head**, **T**, **Lewis**, **D**, **Nov**, **O**, **Shaer**, **O**, **Tzovara**, **A**, **Bobe**, **J**, and **Price Ball**, **M** 2019 Open Humans: A platform for participant-centered research and personal data exploration. *GigaScience*, 8(6). DOI: https://doi.org/10.1093/gigascience/giz076

**Groom**, **Q**, **Weatherdon**, **L**, and **Geijzendorffer**, **I R** 2017 Is citizen science an open science in the case of biodiversity observations? *Journal of Applied Ecology*, 54(2): 612–617. DOI: https://doi.org/10.1111/1365-2664.12767

**Guerrini**, **C J**, **Majumder**, **M A**, **Lewellyn**, **M J**, and **McGuire**, **A L** 2018 Citizen science, public policy. *Science*, 361(6398): 134–136. DOI: https://doi.org/10.1126/science.aar8379

**Hanke**, **G**, **Walvoort**, **D**, **Van Loon**, **W**, **Addamo**, **A**, **Brosich**, **A**, **Del Mar Chaves Montero**, **M**, **Molina Jack**, **M**, **Vinci**, **M**, and **Giorgetti, A.**, **G** 2020 EU Marine Beach Litter Baselines. Luxembourg: Publications Office of the European Union. DOI: https://doi.org/10.2760/16903

**Hansen, J S, Gadegaard, S** and **Holmstrand, K F** 2021 9 things to make citizen science data FAIR. A research librarian's guide. *Technical University of Denmark*, DOI: https://doi.org/10.11583/DTU.12998663

**Heigl**, **F**, **Dörler**, **D**, **Bartar**, **P**, **Brodschneider**, **R**, **Cieslinski**, **M**, **Ernst**, **M**, **Fritz**, **S**, **Krisai-Greilhuber**, **I**, **Hager**, **G**, **Hatlauf**, **J**, **Hecker**, **S**, **Hübner**, **T**, **Kieslinger**, **B**, **Kraker**, **P**, **Krennert**, **T**, **Oberraufner**, **G**, **Paul**, **K T**, **Tiefenthaler**, **B**, **Vignoli**, **M**, **Walter**, **T**, **Würflinger**, **R**, **Zacharias**, **M**, and **Ziegler**, **D** 2018 Quality criteria catalogue for citizen science projects on Österreich forscht. *Zenodo*. DOI: https://doi.org/10.5281/zenodo.1161948

**Higgins**, **C I**, **Williams**, **J**, **Leibovici**, **D G**, **Simonis**, **I**, **Davis**, **M J**, **Muldoon**, **C**, **van Genuchten**, **P**, **O'Hare**, **G**, and **Wiemann**, **S** 2016 Citizen OBservatory WEB (COBWEB):

A Generic Infrastructure Platform to Facilitate the Collection of Citizen Science Data for Environmental Monitoring. *International journal of spatial data infrastructures research*, 11: 20–48. DOI: https://doi.org/10.2902/1725-0463.2018.13.art8

**Holmstrand**, **K F**, **den Boer**, **S P**, **Vlachos**, **E**, **Martínez-Lavanchy**, **P M**, and **Hansen**, **K K** 2019 Research Data Management (eLearning course). DOI: https://doi.org/10.11581/dtu:00000047

**Holmstrand**, **K F**, **Larsen**, **A V**, **Gadegaard**, **S**, **Hansen**, **J S**, **Hansen**, **K K**, and **Thomsen**, **G S** 2020 FAIR data in a Citizen Science project "Fangstjournalen." DOI: https://doi.org/https://doi.org/10.11581/DTU:00000092

**Hunter**, **Jane** and **Hsu**, **C-H** 2015 Formal Acknowledgement of Citizen Scientists' Contributions via Dynamic Data Citations. In: Allen, R B, Hunter, J., and Zeng, M L (eds.), *Digital Libraries: Providing Quality Information. ICADL 2015. Lecture Notes in Computer Science, vol 9469.* Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-27974-9_7

**ICMJE** n.d. International Committee of Medical Journal Editors Recommendations. Defining the Role of Authors and Contributors. *icmje.org*. Available at http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html [Last accessed 12 June 2020].

**Karasmanis**, **S** and **Murphy**, **F** 2014 Emerging roles and collaborations in research support for academic health librarians. In: *Australian Library and Information Association National 2014 Conference, Melbourne, Australia, 18 Sept 2014*. DOI: https://doi.org/10.13140/2.1.2350.7208

**Kennan**, **M A**, **Williamson**, **K**, and **Johanson**, **G** 2012 Wild Data: Collaborative E-Research and University Libraries. *Australian Academic & Research Libraries*, 43(1): 56–79. DOI: https://doi.org/10.1080/00048623.2012.10722254

**Kissling**, **W D**, **Ahumada**, **J A**, **Bowser**, **A**, **Fernandez**, **M**, **Fernández**, **N**, **García**, **E A**, **Guralnick**, **R P**, **Isaac**, **N J B**, **Kelling**, **S**, **Los**, **W**, **McRae**, **L**, **Mihoub**, **J-B**, **Obst**, **M**, **Santamaria**, **M**, **Skidmore**, **A K**, **Williams**, **K J**, **Agosti**, **D**, **Amariles**, **D**, **Arvanitidis**, **C**, **Bastin**, **L**, **De Leo**, **F**, **Egloff**, **W**, **Elith**, **J**, **Hobern**, **D**, **Martin**, **D**, **Pereira**, **H M**, **Pesole**, **G**, **Peterseil**, **J**, **Saarenmaa**, **H**, **Schigel**, **D**, **Schmeller**, **D S**, **Segata**, **N**, **Turak**, **E**, **Uhlir**, **P F**, **Wee**, **B**, and **Hardisty**, **A R** 2018 Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biological Reviews*, 93(1): 600–625. DOI: https://doi.org/10.1111/brv.12359

**Kosmala**, **M**, **Wiggins**, **A**, **Swanson**, **A**, and **Simmons**, **B** 2016 Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14(10): 551–560. DOI: https://doi.org/10.1002/fee.1436

**Miller-Rushing**, **A**, **Primack**, **R**, and **Bonney**, **R** 2012 The history of public participation in ecological research. *Frontiers in Ecology and the Environment*. DOI: https://doi.org/10.1890/110278

**Moher**, **D**, **Liberati**, **A**, **Tetzlaff**, **J**, and **Altman**, **D G** 2009 Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, 6(7): e1000097. DOI: https://doi.org/10.1371/journal.pmed.1000097

**NVK** 2017 Vejledning om forsikring. *nvk.dk*. Available at nvk.dk/emner/forsikring-og-erstatning/vejledning-om-forsikring [Last accessed 1 September 2020].

**Oberle**, **K M**, **Page**, **S A**, **Stanley**, **F K**, and **Goodarzi**, **A A** 2019 A reflection on research ethics and citizen science. *Research Ethics*, 15(3–4): 1–10. DOI: https://doi.org/10.1177/1747016119868900

**OSPP** 2017 Open Science Policy Platform Recommendations. Luxembourg: Publications Office of the European Union. DOI: https://doi.org/doi:10.2777/958647

**Owen**, **R P** and **Parker**, **A J** 2018 Citzen science in environmental protection agencies. In: Hecker, S, Haklay, M, Bowser, A, Makuch, Z, Vogel, J, and Bonn, A (eds.), *Citizen Science - Innovation in Open Science, Society and Policy*. London: UCL Press. pp. 284–300. DOI: https://doi.org/10.14324/111.9781787352339

**Parthenos** How to Manage Data and Metadata in Citizen Science – Parthenos training. Available at https://training.parthenos-project.eu/sample-page/citizen-science-in-the-digital-arts-and-humanities/how-to-manage-data-and-metadata-in-citizen-science/ [Last accessed 20 May 2020].

**Patrick-Lake**, **B** and **Goldsack**, **J C** 2019 Mind the Gap: The Ethics Void Created by the Rise of Citizen Science in Health and Biomedical Research. *The American Journal of Bioethics*, 19(8): 1–2. DOI: https://doi.org/10.1080/15265161.2019.1639389

**Pettibone**, **L**, **Vohland**, **K**, **Bonn**, **A**, **Richter**, **A**, **Bauhus**, **W**, **Behrisch**, **B**, **Borcherding**, **R**, **Brandt**, **M**, **Bry**, **F**, **Dörler**, **D**, **Elbertse**, **I**, **Glöckler**, **F**, **Göbel**, **C**, **Hecker**, **S**, **Heigl**, **F**, **Herdick**, **M**, **Kiefer**, **S**, **Kluttig**, **T**, **Kühn**, **E**, **Kühn**, **K**, **Oldorff**, **S**, **Oswald**, **K**, **Röller**, **O**, **Schefels**, **C**, **Schierenberg**, **A**, **Scholz**, **W**, **Schumann**, **A**, **Sieber**, **A**, **Smolarski**, **R**, **Tochtermann**, **K**, **Wende**, **W**, and **Ziegle**, **D** 2016 Citizen science for all - A guide for citizen science practitioners. Available at https://www.rri-tools.eu/-/citizen-science-for-all-a-guide-for-citizen-science-practitione-1#

**Pulsifer**, **P L**, **Huntington**, **H P**, and **Pecl**, **G T** 2014 Introduction: local and traditional knowledge and data management in the Arctic. *Polar Geography*, 37(1): 1–4. DOI: https://doi.org/https://doi.org/10.1080/1088937X.2014.894591

**Rauber**, **A**, **Asmi**, **A**, **van Uytvanck**, **D**, and **Proell**, **S** 2015 Data Citation of Evolving Data: Recommendations of the Working Group on Data Citation (WGDC). *Zenodo*. DOI: https://doi.org/10.15497/RDA00016

**RDA-CODATA Legal Interoperability Interest Group** 2016 Legal Interoperability of Research Data: Principles and Implementation Guidelines. *Zenodo*. DOI: https://doi.org/10.5281/ZENODO.162241

**Resnik**, **D B**, **Elliott**, **K C**, and **Miller**, **A K** 2015 A framework for addressing ethical issues in citizen science. *Environmental Science & Policy*, 54: 475–481. DOI: https://doi.org/10.1016/j.envsci.2015.05.008

**Riesch**, **H** and **Potter**, **C** 2014 Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Understanding of Science*, 23(1): 107–120. DOI: https://doi.org/10.1177/0963662513497324

**Rothstein**, **M A**, **Wilbanks**, **J T**, and **Brothers**, **K B** 2015 Citizen Science on Your Smartphone: An ELSI Research Agenda. *Journal of Law, Medicine and Ethics*, 43(4): 897–903. DOI: https://doi.org/https://10.1111/jlme.12327

**Runnel**, **V** and **Wijers**, **A** 2019 Improving the detection of collection-based citizen science projects. *Zenodo*. DOI: https://doi.org/10.5281/zenodo.3364519

**Schade**, **S** and **Tsinaraki**, **C** 2016 Survey report: data management in Citizen Science projects. *JRC Technical Reports*. Luxembourg: Publications Office of the European Union. DOI: https://doi.org/10.2788/539115

**Schade**, **S**, **Tsinaraki**, **C**, and **Roglia**, **E** 2017 Scientific data from and for the citizen. *First Monday*, 22(8). DOI: https://doi.org/10.5210/fm.v22i8.7842

**Sheppard**, **S A**, **Wiggins**, **A**, and **Terveen**, **L** 2014 Capturing quality: retaining provenance for curated volunteer monitoring data. In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*. New York, USA: ACM Press. pp. 1234–1245. DOI: https://doi.org/10.1145/2531602.2531689

**Simonis**, **I** 2018 Standardized Information Models to Optimize Exchange, Reusability and Comparability of Citizen Science Data. A Specialization Approach. *International journal of spatial data infrastructures research*, 13: 38–47. DOI: https://doi.org/10.2902/1725-0463.2018.13.art5

**Skov, C** 2021 Database from citizen science project "Fangstjournalen". [Data set]. *Technical University of Denmark*. DOI: https://doi.org/10.11583/DTU.13795928

**Sturm**, **U**, **Schade**, **S**, **Ceccaroni**, **L**, **Gold**, **M**, **Kyba**, **C**, **Claramunt**, **B**, **Haklay**, **M**, **Kasperowski**, **D**, **Albert**, **A**, **Piera**, **J**, **Brier**, **J**, **Kullenberg**, **C**, and **Luna**, **S** 2018 Defining principles for mobile apps and platforms development in citizen science. *Research Ideas and Outcomes*, 4: e23394. DOI: https://doi.org/10.3897/rio.4.e23394

**Syberg**, **K** 2020 data for Mass Experiment [Data set]. *Zenodo.* DOI: https://doi.org/10.5281/zenodo.3886973

**Tauginienė**, **L** 2019 Ethical concerns in citizen science projects and public engagement related research projects. *Ethical Perspectives*, 26(1): 119–134. DOI: https://doi.org/10.2143/EP.26.1.3286291

**Tweddle**, **J C**, **Robinson**, **L D**, **Pocock**, **M J**, **Roy**, **H E**, and **UK Environmental Observation Framework** 2012 *Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK*. London: Natural History Museum. Available at http://www.ceh.ac.uk/sites/default/files/citizenscienceguide.pdf

**UKEOF's Advisory Group** 2013 The principles of planning, collecting and using citizen science data. Advice Note 2. Available at http://www.ukeof.org.uk/documents/DataAdviceNote2.pdf

**US CSA Data and Metadata WG** 2019 PPSR Core Data & Metadata Standards Repository. *GitHub*. Available at https://github.com/CitSciAssoc/DMWG-PPSR-Core [Last accessed 4 August 2020].

**US EPA** 2019 Handbook for Citizen Science Quality Assurance and Documention -Version 1. Washington, DC: United States Environmental Protection Agency. Available at https://www.epa.gov/sites/production/files/2019-03/documents/508_csqapphandbook_3_5_19_mmedits.pdf

**US GSA** Manage Your Data | CitizenScience.gov. *citizenscience.gov*. Available at https://www.citizenscience.gov/toolkit/howto/step4/ [Last accessed 12 May 2020].

**Venturelli**, **P A**, **Hyder**, **K**, and **Skov**, **C** 2017 Angler apps as a source of recreational fisheries data: opportunities, challenges and proposed standards. *Fish and Fisheries*, 18(3): 578–595. DOI: https://doi.org/10.1111/faf.12189

**Wahlberg**, **M** 2020 Harbour porpoises sightings 2019 [Dataset]. *Zenodo*. DOI: https://doi.org/10.5281/zenodo.3661479

**Wang**, **Y**, **Kaplan**, **N**, **Newman**, **G**, and **Scarpino**, **R** 2015 CitSci.org: A New Model for Managing, Documenting, and Sharing Citizen Science Data. *PLOS Biology*, 13(10): e1002280. DOI: https://doi.org/10.1371/journal.pbio.1002280

**Ward-Fear**, **G**, **Pauly**, **G B**, **Vendetti**, **J E**, and **Shine**, **R** 2020 Authorship Protocols Must Change to Credit Citizen Scientists. *Trends in Ecology & Evolution*, 35(3): 187–190. DOI: https://doi.org/10.1016/j.tree.2019.10.007

**Wiggins**, **A**, **Bonney**, **R**, **Graham**, **E**, **Henderson**, **S**, **Kelling**, **S**, **Littauer**, **R**, **Lebuhn**, **G**, **Lotts**, **G**, **Michener**, **W**, **Newman**, **G**, **Russel**, **E**, **Stevenson**, **R**, and **Weltzin**, **J** 2013 Data Management Guide for Public Participation in Scientific Research. *DataOne*. Available at http://safmc.net/wp-content/uploads/2016/06/Wigginsetal2013_DataManagementGuidePPSR.pdf

**Wiggins**, **A** and **Wilbanks**, **J** 2019 The Rise of Citizen Science in Health and Biomedical Research. *The American Journal of Bioethics*, 19(8): 3–14. DOI: https://doi.org/10.1080/15265161.2019.1619859

**Wilkinson**, **M D**, **Dumontier**, **M**, **Aalbersberg**, **Ij J**, **Appleton**, **G**, **Axton**, **M**, **Baak**, **A**, **Blomberg**, **N**, **Boiten**, **J-W**, **da Silva Santos**, **L B**, **Bourne**, **P E**, **Bouwman**, **J**, **Brookes**, **A J**, **Clark**, **T**, **Crosas**, **M**, **Dillo**, **I**, **Dumon**, **O**, **Edmunds**, **S**, **Evelo**, **C T**, **Finkers**, **R**, **Gonzalez-Beltran**, **A**, **Gray**, **A J G**, **Groth**, **P**, **Goble**, **C**, **Grethe**, **J S**, **Heringa**, **J**, **’t Hoen**, **P A .**, **Hooft**, **R**, **Kuhn**, **T**, **Kok**, **R**, **Kok**, **J**, **Lusher**, **S J**, **Martone**, **M E**, **Mons**, **A**, **Packer**, **A L**, **Persson**, **B**, **Rocca-Serra**, **P**, **Roos**, **M**, **van Schaik**, **R**, **Sansone**, **S-A**, **Schultes**, **E**, **Sengstag**, **T**, **Slater**, **T**, **Strawn**, **G**, **Swertz**, **M A**, **Thompson**, **M**, **van der Lei**, **J**, **van Mulligen**, **E**, **Velterop**, **J**, **Waagmeester**, **A**, **Wittenburg**, **P**, **Wolstencroft**, **K**, **Zhao**, **J**, and **Mons**, **B** 2016 The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1): 160018. DOI: https://doi.org/10.1038/sdata.2016.18

**Williams**, **J**, **Chapman**, **C**, **Leibovici**, **D G**, **Lois**, **G**, **Matheus**, **A**, **Oggioni**, **A**, **Schade**, **S**, **See**, **L**, and **van Genuchten**, **P P L** 2018 Maximising the impact and reuse of citizen science data. In: Hecker, S, Haklay, M, Bowser, A, Makuch, Z, Vogel, J, and Bonn, A (eds.), *Citizen Science - Innovation in Open Science, Society and Policy*. London: UCL Press. pp. 321–336. DOI: https://doi.org/10.14324/111.9781787352339

**Wolf**, **M**, **Trejos**, **G**, **Hoeberechts**, **M**, **Flagg**, **R**, **Jenkyns**, **R**, **Morley**, **M**, **Biffard**, **B**, **Kot**,

**M**, **Hogman**, **N**, and **Tomlin**, **M** 2019 Best Practices in Data Management at Ocean Networks Canada: a Citizen Scientist case study. In: *OCEANS 2019 MTS/IEEE SEATTLE*. Ocean Networks Canada, Victoria, Canada: IEEE. pp. 1–6. DOI: https://doi.org/10.23919/OCEANS40490.2019.8962800