Data Management Plan

Port Townsend Marine Science Center
Citizen Science Program

Emily Neal, Citizen Science Educator, AmeriCorps
Jamie Landry, Citizen Science Coordinator (former)
Port Townsend Marine Science Center
February 2014

Edited January 2017, Betsy Carlson, Citizen Science Coordinator (current)

Table of Contents

I. Introduction

This document identifies procedures for working with data as it relates to the Citizen Science Program at the Port Townsend Marine Science Center (PTMSC). It is an **evolving** document to which appropriate amendments should be made. The Data Management Plan's (DMP) aims to identify rigorous methods for each step of data management in PTMSC's Citizen Science projects. The plan ensures data collected by PTMSC volunteers upholds rigorous standards, ensuring the Citizen Science program's credibility with the scientific community. The following steps as well as team members responsible for each step should be thought out prior to the start of the project.

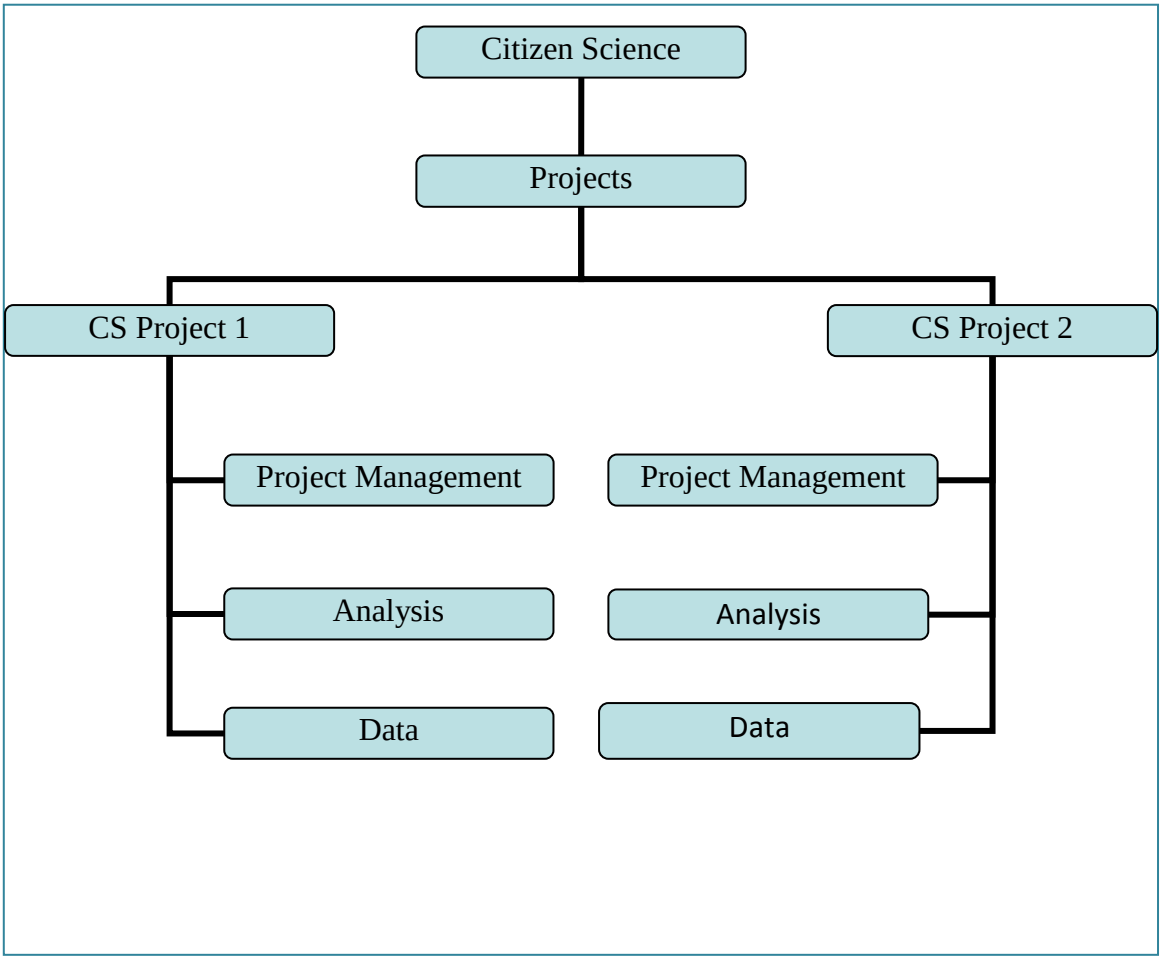
A. Preparing to collect data
B. Data recording
C. Data processing
D. Data analysis
E. Data sharing
F. Data storage

This DMP references several organizations' plans, giving valuable perspective to each aspect/component of data management. It provides an implementation timeline so PTMSC may adopt the plan efficiently. While adopting this plan will require an investment of staff time and resources, these steps will result in reputable and replicable research results, which is of utmost importance to the program. PTMSC believes it is especially important to ensure quality and validity of Citizen Science data because it is often scrutinized more heavily than data collected by trained scientists (Thelen and Thiet).

II. Recommended Plan

There are two important documents that accompany this DMP. They are found in the appendix and in our files: P:\Citizen Science\Data Management Plan, and titled Preliminary Project Report Form and Project Timeline for Data Management. Both documents will help plan for sound data management practices and completing them is a requirement for project supervisors.

All project folders on the PTMSC network drive are organized with the same system. The following diagram defines that system:



A. Preparing to Collect Data

#	Recommendation
A1	Assessing representativeness and comparability
A2	Creating project templates

1. Assessing representativeness and comparability

Proposed Citizen Science (CS) project data's representativeness and comparability must be considered. While it is difficult to quantitatively analyze these elements, the project supervisor should consider the guiding questions listed below, and include their answers in each CS project's preliminary report. This preliminary report should be saved on PTMSC's network drive in the Project Management Folder. Guidelines for the preliminary report are found in P:\Citizen Science\Data Management Plan

i. Representativeness: Will the data be representative of the conditions being sampled? Factors to consider include location, timing, and frequency of sampling. If representative data is unrealistic to collect (for example, it isn't feasible to collect data on private property without permission), this should be noted in the preliminary project report.

ii. Comparability: Will it be possible to compare data collected at various sites within this study? Will it be possible to compare data collected in this study to other studies? In order to strengthen potential comparisons, data parameters collected at each location must remain consistent over time. It should be documented if this isn't possible (for example, the pH meter breaks and pH can no longer be collected). Records of such complications should also be kept in P:\Citizen Science\Projects\Project Name\Project Information (see Appendix A for the recommended filing system).

2. Creating data sheets

Once data selection criteria have been finalized, a field data sheet should be created to ensure the same data is collected at each site and data parameters remain consistent over time. This data sheet should be formatted in a way that will facilitate entry of data into Excel. This forces data considerations to be made before field work is carried out and allows for better document control. It is also possible that field data sheets will be created and managed by a larger research organization for which the data is being collected. For example, the Puget Sound Seabirds Study (PSSS) records data on a worksheet provided by the Seattle Audubon.

B. Data Recording

#	Recommendation
B1	Data collection and field data sheets
B2	Verifying precision and accuracy
B3	Assuring data quality

1. Data collection using field data sheets

Data should be collected on the project's specific field data sheet. "Rite in the Rain®" notebooks or paper should be used if there is a good chance that the records will get wet.

2. Verifying precision and accuracy

Trainings on data collection protocols should be provided by PTMSC staff to project volunteers, ensuring that data collected is precise and accurate. Precision and accuracy of the data collected should be verified in the field by PTMSC staff during "field trainings". Make every effort to ensure the protocol is being precisely followed by all citizen scientists.

i. Precision: If the experiment were repeated multiple times, would there be a low degree of distinction between observed values? Discrepancies in values can be caused by equipment or collection errors. Those collecting data should strive to make it as precise as possible.

ii. Accuracy: With how much confidence is the data collected reflective of the true value? How close is the result to the expected value? Data should be as accurate as possible.

3. Assuring data quality

Data quality checks should be performed by PTMSC staff throughout the project. Possible techniques listed in the table below are taken from Crall et. al's article on 'Improving and integrating data on invasive species collected by citizen scientists' and the EPA's Water Quality's 'Managing Volunteer Data' website (Crall et al; "Managing Volunteer Data").

Table 1. Comparison of Data Quality Checks

Possible Data Quality Checks	Advantages	Disadvantages
Volunteer training	All volunteers receive the same instructions	Requires significant staff time to plan and coordinate; volunteers must all be available
Expert validation of species identification	Species identification is more accurate	Requires hiring of an expert
Inspection of collected field data for completeness and validity by staff member	Flags potentially incorrect data and identifies volunteers who may need extra guidance	Requires staff time
Discarding suspect data	Removes data that is likely inaccurate	May remove accurate outliers

PTMSC currently implements several data quality checks. For example, the SoundToxins program includes semiannual volunteer trainings by PTMSC staff and annual program training by professionals at NOAA. Additionally, SoundToxins volunteers take photographs of species they are not able to identify and send those to experts at NOAA for verification. These data quality checks are an example of effective and appropriate examples of data quality monitoring techniques. Each CS project at PTMSC should be reviewed to determine appropriate data quality checks. It should be noted that according to the National Parks DMP, most quality control comes through defining and enforcing standards (Brumm).

C. Data Processing

#	Recommendation
C1	Uploading data to an appropriate scientific database
C2	Selecting data to upload to PTMSC database
C3	Uploading data to a PTMSC database

1. Selecting data to upload to PTMSC database

PTMSC should store valuable and long-term data on an internal, electronic database so the organization may study long-term trends. Deciding which data is appropriate to store in PTMSC's database must be determined on a case-by-case basis, weighing time needed to upload the data against the potential benefit of studying it more in depth. Record of this decision should be stored in the project's Preliminary Project Report, stored in the Project Management folder. Emphasis should be placed on electronically storing quantitative data rather than qualitative, as trends are more apparent in quantitative. Frequently collected data should also be given priority as larger sample sizes produce more statistically sound analyses.

2. Uploading data to an appropriate scientific database

Handwritten field data should be entered in an appropriate electronic database in a timely manner. It is possible that some PTMSC projects will not require this step. For example, samples from the Mussel/PSP Sampling Project are sent directly to a laboratory for analysis and no ambient environmental data is collected during the process.

Table 1. Data Quality Checks During Data Entry

Data Quality Checks To Be Completed by Staff Member	Outcomes	Special Considerations
Inspection of collected field data for completeness and validity by staff member	Flags potentially invalid data and identifies volunteers who may need extra guidance	Requires staff time
Discarding suspect data	Removes data that is likely inaccurate, be cautious with this step	May remove accurate outliers, only take this step when it is certain that the entry was inappropriately recorded

3. Uploading data to a PTMSC database

Microsoft Excel workbooks should be used to digitalize CS project data. The program is commonly known and useful for manipulating data, analyzing trends of over time, and representing it in charts and graphs. It is relatively easy to import data among different versions of Excel in case PTMSC chooses to update its software. Additionally, data entered in Excel is compatible with a wide range of other data software.

These Excel spreadsheets should be stored on PTMSC's internal network drive so they may be accessed by any staff but the public may not alter them. They should be stored in the 'Data' folder appropriate for each project (see Appendix A). It is possible each project will need more than one Excel workbook as this depends on the specific data collected. For example, different results from the same marine mammal stranding network necropsy could be divided into different columns, tabs, or even workbooks within Excel. Determine on a case-by-case basis.

D. Data Analysis

#	Recommendation
D1	Data completeness
D2	Analyzing data

1. Data completeness

Before analyzing project data, its completeness should be considered. This is a percentage measuring the anticipated usable data versus what was actually collected. Ideally, well-prepared data collection will be 100% complete. However, this allows for no error in the case of broken instruments or sampling error. If data is less than 90% complete, the representativeness and comparability of the collected data should be reconsidered (Hofferkamp). A separate document form in the analysis folder should be created.

2. Analyzing data

Priority should be given to processing data that is quantitative and extensive (i.e. data that was consistently taken in the same location over the course of months/years). Examples include studying pH or water quality parameters at a specific location over long periods of time.

CS staff should verify that analysis pursued by PTMSC avoids redundancy. For example, the PSSS website posts a graph of the total number of each species of bird found at the Fort Worden Site (run by PTMSC); therefore, it would not be an appropriate use of PTMSC staff time to graph that information.

E. Data Sharing

#	Recommendation
E1	Sharing results with scientific community via online databases and publications
E2	Sharing results with CS volunteers via posting in the lab, trainings, and possibly emails
E3	Sharing results with visiting public via posting in the lab and encouraging docents to point out the posts
E4	Sharing results with public online on PTMSC website
E5	Considering data's sensitivity before publishing

"The main reason a Data Management Plan is required, is for you to think about how you prepare (manage) your data for sharing and describe how you will actively share your data with non-group members after the project is completed" (DMP Tool).

There is a spectrum of groups with whom PTMSC should share its collected and analyzed data. It is important to further the collective knowledge of the scientific community in and around the Salish Sea. Sharing results with CS volunteers and educating the public about ongoing research at PTMSC is vital. All this can be done through prioritizing data sharing (Insel). See the Citizen Science Program Evaluation Report for more information on how to increase our program's transparency, sharing, and networking efforts.

1. Considering data's sensitivity before publishing

While little of the data collected by PTMSC will result in the devaluation of private property or offend visitors with its implications, PTMSC should be aware of this possibility and avoid publicizing sensitive information. For example, it is feasible that a visiting scallop fishery owner would be upset to find data showing how waters near his business have quantifiable levels of *pseudo-nitzschia*, even if that level isn't high enough to require him to shut down. Also important to data sensitivity is checking that PTMSC has legal rights to any data it publicizes. The SoundToxins websites includes the clause "I shall contact NOAA prior to publication of any analyses using data from this site to review acceptable use of the data" (SoundToxins.org Volunteer Data Entry). Any known information to this effect should be included in the initial project report and titled Data Sensitivity, saved under the 'Project Information' Folder (see Appendix A).

2. Sharing results with scientific community via online databases and publications

PTMSC has succeeded in sharing data with the greater scientific community. Because many projects are coordinated by larger research organizations, PTMSC has contributed results to their online databases, allowing results to be compared to those from other sampling teams in the region.

Some CS projects such as the Plastic Ingestion Study were run independently by PTMSC. This study was published in a peer-reviewed journal, a different and equally valuable method of sharing data with the scientific community. PTMSC should continue to share its findings with other regional scientists, educators, and resource managers.

Methods of sharing data with the scientific community:

- Upload data to online scientific databases of larger research organizations
- Publish data (and analyses) in peer-reviewed scientific journals
- Posting results on PTMSC's Citizen Science webpage Publishing semiannual results in newsletter "PTMSC Citizen Science Update"

3. Sharing results with CS volunteers via posting in the Discovery Lab, trainings, and possibly emails

PTMSC should ascertain that results are shared with CS volunteers whenever possible. This validates the usefulness of the volunteers' work (Rotman, Preece et al). This is especially important for analysis we receive from organizations with whom we collaborate. For instance, summaries of SoundToxins data are available upon request from NOAA and the PSSS website posts graphs of bird counts by site. PTMSC also received a summary of World Ocean Day results. Recommended methods of sharing this information include posting it in the Foss Maritime Discovery Lab, displaying and explaining it during a volunteer training, or even emailing it to interested volunteers and posting on the Citizen Science webpage and blog.

4. Sharing results with visiting public via the Discovery Lab and encouraging docents to point out the posts

PTMSC seeks to "Inspire Conservation of the Salish Sea" and therefore has a duty to share research results with the general public. Posting CS results in the lab will help (this space is open to the public during visitor hours), but only if visitors enter the lab and are encouraged to look around. PTMSC staff should continue to make the space more inviting to visitors and find ways to engage the public. Staff should also ensure that docents are trained on ways to interpret research results, encourage visitors to step into the Discovery Lab, and engage with Citizen Science interactive in the exhibit/Lab spaces.

5. Sharing results with public online on PTMSC website

An important and long-term goal for PTMSC to consider is making appropriate research results available online. A successful model of this can be found at the H. J. Andrews Experimental Forest. In their program, the public can access archives of research results from on site projects. Examples of accessible data include historical and current weather conditions ("HJ Andrews Experimental Forest Long Term Ecological Research"). Because of their effort to make this data public, affiliated scientists from Oregon State University can supplement current research with data from H.J. Andrews.

Publishing PTMSC data would also provide local high school students the opportunity to incorporate local scientific research into school projects. Additionally, having graphs or other analytical representations available would improve public outreach and future online education materials. Public outreach for the CS program would result in raised awareness of scientific issues and processes, as well as help PTMSC gain more respect as a scientific institution. More publicity through PTMSC's website and exhibits would allow the casual visitor to learn more about the CS program at PTMSC.

Possible methods of sharing with public:

- PTMSC website

- PTMSC Facebook page/Twitter account
- PTMSC Blog
- PTMSC E-news
- Local newspapers (PT Leader, etc.)
- PTMSC Citizen Science Newsletter

F. Data Storage

#	Recommendation
F1	Storing virtual data indefinitely
F2	Long term data storage

1. Storing virtual data indefinitely

It is important to properly store data in order to facilitate efficient data retrieval. Electronic data should be kept in its original Microsoft Excel workbook on PTMSC's internal network drive. Digital copies should be kept indefinitely as Excel documents don't use excessive computer memory and are able to store vast amounts if information. It may become necessary to archive old data that can no longer be used for relevant analysis. This could be done within the same project folder (P:\Citizen Science\Project Name\Data) under a subfolder titled 'Archived Data'; alternatively, an 'Archived Data' folder could be made directly under the project folder (P:\Citizen Science\Project Name\Archived Data).

Maintaining a backup digital copy of the data is required as preventative measures for a circumstance in which data in the PTMSC internal drive is lost. This back up should be carried out once a year following the New Year. The data included in the back up should include projects where data has been collected over the past year. Any data file that is backed up should be recorded by year in the "data_backup_list", located in the DMP file in the PMSC internal drive.

2. Long term data storage

It is also necessary to store original hard copies of field data sheets. This ensures researchers questioning the validity of transcribed data the ability to reference original field data sheets. Original field data sheets may have notes from the recorder which were not able to be added to the computer database. It also provides PTMSC the ability to view data that was collected but not analyzed. Ideally, hard copies of data should be kept indefinitely. However, as this may take up valuable space, it may not be possible and data which is $\geq$15 years may be considered for disposal. Recent data (age < 1 year) should be stored in an accessible cupboard in the PTMSC Foss Maritime Discovery Lab. Older data may be stored in a less accessible cabinet in the Discovery Lab or the PTMSC shop.

III. Proposed Timeline

This DMP is an evolving document and steps will be taken to continuously update it after its publication in the winter of 2013. Under the direction of Citizen Science Coordinator, Jamie Landry, there will be ongoing updates. It is our hope that this DMP will serve as a resource to other organizations looking to produce similar documents and allow PTMSC to further its contributions to the scientific community.

A. 2013-2014 AmeriCorps season

Action to take
Continue sharing results with volunteers and public in a timely manner
Determine archived data to be computerized
Computerize archived data
Analyze and share long term trends

The AmeriCorps member hired as Citizen Science Lab Coordinator for the 2013-2014 seasons should continue to share analyses of collected data with volunteers and visiting public by posting in the Discovery Lab.

It is necessary to determine which data collected would be suitable to digitalize and analyze. Suitable examples include studying pH or water temperature at a specific location over many years – this data should have been collected by SoundToxins and possibly other projects as well. Analyzing the number of *pseudo-nitzschia* over time, either at a specific site or added between the sites studied, would also be useful. This step should be completed in fall 2013 and digitalized over winter 2014 by the CS AmeriCorps member. Volunteers can also help with this task. Ideally, preliminary analysis will be completed and posted by the reopening of the Marine Exhibit portion of PTMSC in spring 2014.

B. Long Term

Action to take	Timeframe
Determine relevant data/analyses to publish on PTMSC's website	2-3years
Collect relevant data/analyses, format it, and determine feasibility of publishing it on PTMSC's website	3-5 years
Publish data/analyses on PTMSC's website, update periodically	5-10 years

PTMSC should continuously work to post collected data and analyses on its website. Posting data online would allow local high school students to do projects which incorporate locally collected research data. This would help PTMSC connect with and maintain channels of communication with high school students and administrators. Posting analyses online would lead to greater awareness of CS projects at PTMSC. Relevant data/analyses should be chosen for publication and formatted before publishing online and contracted personnel should maintain that portion of the PTMSC website.

IV. Future Resources

A key purpose of this DMP is to assist PTMSC staff in writing subsequent DMPs for future CS projects and grant reports. Resources referenced throughout this paper will provide useful for those cases.

The National Park Service has published online a very comprehensive (pages ≈ 100) DMP for their Great Plains Inventory and Monitoring Network. It provides a well reviewed and established reference for PTMSC's future data management endeavors (Brumm).

DMPs are often required as a reporting requirement by funders and grant managers. The DMP Tool website is specifically designed to help users fill out a DMP for a specific grant (DMP Tool). They have a variety of formats to choose from depending on the specific grant-giving institution and type of grant. There is also a miscellaneous grant section called 'NSF – generic.' Each format gives you specific questions to answer under each section, which allows the user to properly address relevant points without including unnecessary information. This online tool should be used when specialized DMPs are required of PTMSC. The PTMSC username/password can be found under P:\Citizen Science\Data Management Plan\Resources.

V. Conclusion

It is often the case that specific grant-giving companies require a DMP for project proposals. This document will aide in completing those DMPs, as will the references listed under the 'Future Resources' section of this report. This document is intended to guide Citizen Science projects at PTMSC so all projects will follow defined standards, improving credibility and continuity of CS data collected at PTMSC. Lastly, this document serves as a guide for other organizations aiming to write their own DMP.

VI. References

Brumm, J. 2009. Data Management Plan: Northern Great Plains Inventory and Monitoring Network. National Park Service Northern Great Plains Inventory and Monitoring Network Report. NGPN/2009.

Crall, A. et al. "Improving and integrating data on invasive species collected by citizen scientists." *Biological Invasions*. 07 2010. Print.

DMP Tool: Guidance and Resources for your Data Management Plan. University of California Curation Center, 2010. Web. < https://dmp.cdlib.org/> 06 Dec. 2012.

Hofferkamp, J (Ed). Standards for Cancer Registries Volume III: Standards for Completeness, Quality, Analysis, Management, Security and Confidentiality of Data. Springfield (IL): North American Association of Central Cancer Registries, August 2008.

Insel, Thomas. "Time Matters - Why We Care So Much About Data Sharing." *National Institute of Mental Health*. N.p., 21 May 2012. Web. 23 Dec 2012. <http://www.nimh.nih.gov/about/director/2012/time-matters-why-we-care-so-much-about-data-sharing.shtml>.

Port Townsend Marine Science Center: Traveler Reviews. Trip Advisor, 2011. Web. < http://www.tripadvisor.com/ShowUserReviews-g58687-d1557428-r121290703-Port_Townsend_Marine_Science_Center-Port_Townsend_Washington.html#REVIEWS> 02 Dec. 2012.

Rotman, Dana, Jennifer Preece, et al. "Dynamic Changes in Motivation in Collaborative Citizen-Science Projects." Web. 14 Jan. 2013. <http://www.danarotman.com/wp-content/uploads/2012/01/CSCW_final_for-publication.pdf>.

"Search the Andrews Data Catalog." *HJ Andrews Experimental Forest Long Term Ecological Research*. Web. <http://andrewsforest.oregonstate.edu/lter/data.cfm?frameURL=http://andlter.forestry.oregonstate.edu/lter meta/ltersearch/DataCatalog.aspx?topnav=8> 28 Nov 2012.

SoundToxins.org Volunteer Data Entry. SoundToxins.org. Web. <https://www.webapps.nwfsc.noaa.gov/apex_bolt/f?p=264:1> 05 Dec. 2012.

Thelen, Brett, and Rachel Thiet. "Cultivating connection: Incorporating meaningful citizen science into Cape Cod National Seashore's estuarine research and monitoring programs." *Park Science: Integrating Research and Resource Management in the National Parks*. 10 2012. Web. 27 Dec. 2012. <http://www.nature.nps.gov/ParkScience/index.cfm?ArticleID=236&Page=1>.

United States. Environmental Protection Agency. *Managing Volunteer Data*. Web. <http://water.epa.gov/type/rsl/monitoring/vms61.cfm>.

VII. Appendices

A. Proposed Timeline

2013-2014 AmeriCorps Season	Timeframe
Continue sharing results with volunteers and public in a timely manner	Ongoing
Computerize data – sound toxins mainly, see implementation document for more projects to work on	Fall 2013
Write Online data sharing section in DMP	Winter 2014
Analyze and share long term trends	Spring 2014
Long Term	**Timeframe**
Determine relevant data/analyses to publish on PTMSC's website	2-3years
Collect relevant data/analyses, format it, and determine feasibility of publishing it on PTMSC's website	3-5 years
Publish data/analyses on PTMSC's website, update periodically	5-10 years

B. Preliminary Project Report example and form

PORT TOWNSEND MARINE SCIENCE CENTER

(example) Preliminary Project Report

For Data Management

The elements of the PTMSC Data Management Plan that are to be completed ***prior*** to the start of a Citizen Science project are summarized in the table on the next page. Complete each element and add comments to the "steps taken/notes" section of the form. Examples of how to complete the preliminary project report are below.

Sections of DMP to be completed *prior* to starting project	**Steps taken/notes**	**Date completed**
Part A1 – Preparing to collect data; Assessing representativeness and comparability	*Example: It was identified during the project planning process that sampling 2 times a year does not provide a sufficient representation of seasonal variation, considering there are 4 distinct seasons of weather patterns in the Port Townsend area. In order to address this, we altered the sampling protocol to include 4 samplings each year.*	1/1/2014
Part E – Data sharing	Which methods of data sharing will the project use? *Example: This project lends itself well to posting quarterly sampling results on Facebook and the PTMSC blog. Annually, a summary report will be shared in the volunteer e-news and on the PTMSC CS webpage.*	1/1/2014

Preliminary Project Report

For Data Management

Project Name:

Project Supervisor/Prepared By:

Sections of DMP to be completed *prior* to starting project	**Steps taken/notes**	**Date completed**
Part A – Preparing to collect data; Assessing representativeness and comparability		
Part A – Preparing to collect data; Creating data sheets		
Part B – Data recording; Verifying precision and accuracy	What steps will be taken to verify precision and accuracy during data collection?	
Part B – Data recording; Assuring data quality	Which methods of data quality checks will be used during data recording?	

Part C – Data processing; Uploading to appropriate database	What non-PTMSC databases will data be uploaded to?	
	What data quality checks will be used during data processing?	
Part C – Data processing; Uploading to PTMSC database	What format/program will data be entered in?	
Part D – Check for anomalies	What are the potential sources and how can protocols be adapted/improved?	
Part F – Data sharing	Which methods of data sharing will the project use?	
Part F – Data sharing; Consider data's sensitivity	Any special considerations?	
Part G – Data storage	Where and how will data be stored?	

PORT TOWNSEND MARINE SCIENCE CENTER AT FORT WORDEN STATE PARK

Data Management Timeline – Citizen Science

Project Name: _____

Project Supervisor: _____

***use timeline as checklist for completing aspects of data management according to the PTMSC Data Management Plan (DMP)**

Part A – Preparing to collect data

- ☐ Assess representativeness and comparability
- ☐ Determine protocol, create data sheets
- ☐ Assure data quality
- ☐ Train volunteers on protocol

Part C- Data processing/entry

- ☐ Select data to upload
- ☐ Upload data

Part F- Data Sharing

- ☐ Complete project's data sharing profile
- ☐ Complete project's metadata profile
- ☐ Share with scientific community, volunteers, public

Project Selected | **Immediately** | **In the field** | **Ongoing** | **In conclusion**

Preliminary project report

Part B – Data recording

- ☐ Use field data sheets
- ☐ Verify precision and accuracy
- ☐ Assure data quality in field

Part D –Assuring data quality

- ☐ Check for data completeness
- ☐ Check for field errors
- ☐ Check for anomalies

Part E – Data analysis

- ☐ Data completeness
- ☐ Analyze data
- ☐ Identify overall trends

Part G – Data storage

- ☐ Store electronic data
- ☐ Long term data