

Coastal and Marine Science Center Data Management Plan

Project Information

Project Name:

This should be the same name entered into BASIS. The name must be unique.

1954 Bathythermograph Data Rescue

BASIS Number:

This will allow future entries in Compass to use the DMP and update the DMP

(*Personally identifiable information has been removed from this example)

Principal Investigator(s):

John Smith

Collaborating Agencies/Organizations:

Great Lakes Science Center
Fort Collins Science Center

Collaborator Contact Info:

(*Personally identifiable information has been removed from this example)

Summary of Project:

This should be the same as in BASIS

This data represents a small subset of a much larger collection of USGS bathythermograph slides and images from the Great Lakes region of the U.S. and Canada, focusing on bathythermograph data (temperature at depth) from 1954. To produce this data release, 1330 bathythermograph photographic prints were scanned and software was used to extract temperature and depth data from the scanned images. Resulting data was quality controlled for accuracy and published in accordance with USGS Fundamental Science Practices for data releases. The purpose of digitizing this subset of images is to provide better access to the data, and preserve information at risk of loss. Additionally, data regarding the resources and time required to preserve and release this legacy dataset will be used to better evaluate and prioritize future USGS legacy data preservation efforts, including preservation of the remaining USGS bathythermograph collection.

Project Start Date:

This should be the same as in BASIS.

(2016-08-17)

click in the text field to set the date

Project End Date:

This should be the same as in BASIS.

(2017-01-20)

click in the text field to set the date

Information Specialist:

This could be the PI or a designee who has overall responsibility for data acquisition, processing, quality control, documentation, and preservation. This person is likely to be helped by data specialists and oversee their work.

(*Personally identifiable information has been removed for this example)

Plan and Acquire

Will this project collect new data?

Yes

No

Will this project use existing USGS data?

- Yes
- No
- Unknown

If you answered 'Yes' above, where did existing dataset(s) come from?

Original temperature, depth, and cruise data were gathered in the field by the Bureau of Commercial Fisheries in the format of mechanical bathythermograph-etched slides. Over time this data has been transferred to the USGS Great Lakes Science Center. These data have been converted to photographic film and prints and will be stored and accessed through the DaR preservationist at the USGS Fort Collins Science Center.

Is that data published or publicly available?

- Yes
- No
- Unknown

Will this project use contracted or donated datasets (data from external sources)?

- Yes
- No
- Unknown

If you answered 'Yes' above, where did contracted or donated dataset(s) come from?

Will you be able to store a local copy of the contracted/donated data?

- Yes
- No
- Unknown

Estimated volume of all the data that will be collected, acquired, and/or generated:

Approximately 1330 3x5 inch photographs will be scanned, creating roughly 30 GB of digital

data.

Temperature and depth data will be interpreted and entered into a database that is expected to be roughly 100 MB.

MB or TB?

MB

Will the project need to buy any new software, licenses, or IT equipment to collect or process the data?

- Yes
- No
- Unknown

If you answered 'Yes' above, please explain below.

How many payperiods have been requested for project staff to assist with these tasks?

- | | | |
|--|---------|-------|
| 1. Project planning: | 0.5 pp | <hr/> |
| 2. Data acquisition: | 0.25 pp | <hr/> |
| 3. Data processing: | 2 pp | <hr/> |
| 4. Data analysis: | 0 pp | <hr/> |
| 5. Data backup and archiving: | 1 pp | <hr/> |
| 6. Publication and data release: | 3 pp | <hr/> |
| 7. Prepare metadata and documentation: | 1 pp | <hr/> |
| 8. QA/QC: | 2 pp | <hr/> |

Describe and Manage Quality

Are you publishing or archiving data in a proprietary or non-proprietary format?

- Proprietary
- Non-proprietary
- Both

Are you archiving data in a digital or non-digital form?

- Digital
- Non-digital
- Both

Are file naming conventions going to be used that help describe the data?

- Yes
- No

How will the raw, unpublished data files be documented for future usefulness?

Bathythermograph gold/smoke plated etched glass plates are considered the master files for this project. However, bathythermograph etched slides can only be read with a calibrated bathythermograph, therefore, a secondary format, photographic prints, are being used for this project. These are, for all intents and purposes, are identical to the master files. For the 2016 Data at Risk project, legacy data in the form of photographic prints of project bathythermographs was provided to the DaR project for preservation and release purposes only. Prints were made from the slides shortly after original data acquisition and are therefore identical, they are less fragile to handle than glass slides and they are easily interpreted in digital format. All photos will be returned to the GLSC once data has been published. Documentation of the photographs will be included in the final metadata record for the dataset.

Who is going to be responsible for producing your metadata in a FGDC-endorsed standard?

Metadata will be generated by Project Staff.

Describe planned QA/QC procedures during data acquisition and processing.

- FGDC CSDGM Metadata standard
- Federal Agencies Digitization Guidelines Initiative (draft 2015)
- USGS bathythermograph users manual

- All printed photos are assumed correct in their printed format.
- Temperature, depth and cruise data that is collected from the scanned prints will be reviewed and verified by an additional preservation team member at a rate of 100% at the beginning to ensure data collection methods are being consistently applied and reducing over time as consistency is verified as preservation staff gain experience. We will tentatively follow the schedule below, adjusting as reviews/evaluations
- dictate.

| Print Image | %Review/Verify |
|-------------|----------------|
| 1-100 | 100% |
| 101-700 | 75% |
| 601-900 | 50% |
| 901-1330 | 25% |

Which of the following will be included with your data?

- Supporting documents (i.e. fieldbooks)
- Standardized logs (i.e. FACS logs)
- Standard operating procedure

Backup/Secure and Preservation

See your local IT group for help on these questions.

Where does the project plan to store copies of research files and data to ensure their safety?

The database will be maintained in Google Doc's, The original bathythermograph-etched glass slides are stored at the USGS Great Lakes Science Center, and the photo prints of bathythermograms are being stored in a secured, climate stable room at the USGS Fort Collins Science Center (Rm 1113) during preservation and publication, after which, they will be returned to their permanent storage at the GLSC.

Where are hardcopy notebooks, disk copies of data, and physical samples going to be physically stored?

- The original bathythermograph-etched glass slides are stored at the USGS Great Lakes Science Center.
- Photo prints of bathythermograms are being stored in a secured, climate stable room at the

USGS Fort Collins Science Center (Rm 1113) during preservation and publication, after which, they will be returned to their permanent storage at the GLSC.

How many copies of data will be maintained and how will they be kept synchronized?

- If error is discovered in the operational copy of the gDocs spreadsheet we will revert to an earlier version of the spreadsheet using the default gDocs “See Revision History” tool.
 - If the operational copy of the gDocs spreadsheet is lost or damaged beyond the limits of reverting to an earlier version, the operational copy will be replaced with the last weekly backup. Weekly backups occur at the beginning of each week before any data entry has been started. Weekly backups are stored in the Bathythermograph data’s Sciencebase item under “Database Backups.”
-

Are the data and backups going to be stored in multiple places and on different media types to protect against a single-point failure?

- Yes
 No

Who will be responsible for archiving project data on the server?

(*Personally identifiable information has been removed for this example)

Will there be an easily accessible index that documents where all archived data are stored and how they can be accessed?

There will be if you use Compass.

- Yes
 No
 Not sure

If your data is not on the SPCMISC project server, how often will backups be performed?

Weekly.

Who will be responsible for the backups?

This will be IT, if you store data on SPCMSC project server.

All data stored on a server (office shares such as "Shared", user shares, or project shares) and are all backed up by WARC IT staff. Full backup to tape once a month and incremental to disk every weekday night. Backup tapes are kept for 6 months before reuse.

How long are backups going to be kept?

One month.

What will the final format of the data be? Check all that apply.

- ArcASCII
- AVI
- CSV
- DOC
- DOCX
- JPEG
- MAT
- MOV
- MPEG-4
- netCDF
- PDF
- SEGY
- SHP
- TIFF
- TXT
- XLS
- XLSX
- XML

List other formats here.

Is that format appropriate for long-term preservation (i.e. non-proprietary, machine readable)?

- Yes
- No
- Not sure

Release, Publish, and Share

What methods will be used to improve access to data after it is released?

Data will be available online through ScienceBase and the metadata record will be accessible via USGS Science Data Catalog and data.gov. Web services will be made available as Web Map Service (WMS) or JSON.

What methods will be used to publish data interpretations and/or new methods?

No data interpretation will be performed. New methods are not being used/developed for this project.

How is non-interpretive data going to be released?

Non-interpretive data will be released through ScienceBase as a data release.

Will there be a webpage associated with this project?

- Yes
- No

If there is an existing web page, enter the URL below.

Describe project policies for sharing data with internal and external sources before publication.

Internal (USGS) data sharing is allowed. External data sharing will require a data sharing agreement if prior to publication.

Will there be any access or use restrictions on the data (e.g. sensitive data, restricted data, etc.)?

- Yes
- No

If you answered 'Yes' above, please explain.

Is there a deadline for sharing your data as required by the funding agency?

- Yes
- No
- Unknown

If you answered 'Yes' above, what is the deadline?

2017-01-20

Submit