**Data Sharing Plan**

1. **Data Management Plan Justification**

The proposed project will result in the acquisition and temporary storage of between 20 and 30 tbytes of global satellite SST files in netcdf and hdf format, downloaded from the daily global NOAA pathfinder V3.5 SST data set between 1984-2014 at <ftp://ftp.nodc.noaa.gov/pub/data.nodc/pathfinder/Version5.3> and the MODIS LST or SST daily data sets downloaded from one of the many sources listed at: <https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod11a1_v006>

Or <https://modis.gsfc.nasa.gov/data/dataprod/mod28.php>

The derived climatological data sets and associated time series for each of the several hundred lakes that will be processed as part of this study will be permanently archived. The research group that Dr. Crosman is part of at the University of Utah has extensive experience managing large data archives obtained from real-time monitoring sites as well as field programs. For example, MesoWest software (<http://mesowest.utah.edu>) accesses, archives, and makes available environmental observations from over 40,000 surface stations (<http://mesowest.org/api>). We have also recently started a publically-accessible archive for the High Resolution Rapid Refresh Model <http://home.chpc.utah.edu/~u0553130/Brian_Blaylock/cgi-bin/hrrr_download.cgi>

**II. Project Data Management**

**1. Environmental Data types**

*-NOAA pathfinder V3.5 SST*: These data files (in netcdf format) between 1984-2014 will be downloaded each day. These daily files (daytime and nighttime are both produced daily) are each of ~37 Mb is size. Thus, calculating ~37 Mb per day \* 365 days per year \* ~30 years (1985-2014) results in ~1 tbyte of data.

*-MODIS LST or SST*: Depending on how many global swaths of higher-resolution MODIS imagery is downloaded, these files will result in up to 4.8 Gb of data per day \* 365 days per year \* ~17 years results in ~27 tbytes. Because we do not need to extract all global regions, this is an upper estimate and the expected size would likely be below this, and if needed we will sub-set lake regions and not store all of the large files.

*-In Situ Verification Data Sets:* These data sets are small and generally consist of csv, text, or excel spreadsheet files and are not expected to result in any notable storage requirements.

*-Ancillary data sets:* These will include data sets to help aid in the quality control of the LSWT data sets, including the Inland water dataset for distance-to-land, distance-to-water, water-body identifier and lake-centre co-ordinates (Carrea et al. 2015). The total storage requirements of these data sets is expected to be less than 1 tbyte.

**2. Quality Assurance**

*NOAA pathfinder V3.5 SST and MODIS SST/LST*: These data are widely used and provided in an output format (netcdf and hdf) that insures consistency in the fields.

*Verification Data Sets:* All *in situ* observations obtained through the MesoWest or Sharma et al. (2015) archive are processed and will be subjected to quality control procedures upon receipt.

**3. Data Access**

Almost all of the information we will use in our analysis is in the public domain. Only in situ lake validation data sets in original format obtained from individual researchers and lake environmental managers would require permission of the agencies who supplied that information. All of the research to be completed will be undertaken within the secure computing environments of the research teams at the University of Utah and University of Alaska that have appropriate security measures in place to limit direct access to the project team. As the project progresses, processed information will be accessible via a web server at the University of Utah for other members of the research team and anyone interested in the research to help in the evaluation of the project results.

**4. Storage and Backup**

Procedures are already in place to manage the large data sets received and archived at the University of Utah. The PIs research group has over 100 tbytes of disk storage available in the CHPC computing system, with approximately 15 tbytes currently available. As part of this proposal an additional 30 tbytes of disk storage for this project will be purchased by CHPC, allowing for a total storage for this project of up to 45 tbytes. CHPC provides a robotic tape system that is capable of hosting critical data required for our study and the Principal Investigator has sufficient numbers of tapes to store a backup of all data resources during the project.

**III. Long-Term Data Management**

1. **Metadata**

Our metadata documents will be included with the NOAA National Centers for Environmental Information (NCEI) data repository described below prior to completion of the project conforming to appropriate standards specified by NOAA. Any publications that result from this research and papers describing the data sets will be submitted to the NOAA Institutional Repository after acceptance. We will update earlier metadata provided to NOAA as needed: repository name, URL, and repository-assigned identifier.

1. **Data Repository**

The data repository for the derived climatological lake temperature fields will be the NOAA NCEI. We submitted a contingent request for archiving data in the NCEI in July 2017 and were approved prior to submitting this proposal. We have experience undertaking research that culminates in the underlying data being archived in a data repository (doi:http://dx.doi.org/10.5065/D6028PRS) and having that data described in a peer-reviewed data journal (e.g., Jacques et al. 2016). The Marriott Library at the University of Utah is in the process of developing a data repository designed to meet the needs of researchers who intend to archive data relied upon to complete sponsored research. We have been in contact with the Library’s data repository team to assess whether the resources will be available at the completion of our project to archive the downloaded raw Pathfinder and MODIS data evaluated in this study (this data is not being archived at NOAA NCEI, only the quality controlled derived climatology). Because of the potential high interest in the data for further research, we will pursue describing the resulting data set in a publication in a data science journal, such as that by Sharma et al. (2015).

1. **Data Access**

All of the data permanently archived at NOAA NCEI or the Marriott Library data repository will be classified as “open access” and will be publically available free of charge by within the two year time window required by NOAA.