PacMARS Metadata Template (preliminary):

The metadata template is based on the ISO 19115-2 metadata standard.

What does this data set describe?

Title: Permanent open surface water areas in Nenets Autonomous Okrug, Russian Federation

Abstract: Overview / Item Description / Description (Abstract)

Topic Category: Overview / Topics & Keywords / Topic Categories

Purpose: Overview / Item Description / Summary (Purpose)

a) How should this data set be cited? Overview / Citation / Other Details? Overview / Item Description / Credits?

Author: Overview / Citation Contacts / Resource Citation Contacts (specify Author as Role)

Philippe Amstislavski[1]
Leonid Zubov[3]
Herman Chen, MBA[1]
Pietro Ceccato[2]
Jean-Francois Pekel[4]
Jeremy Weedon[1]

1-State University of New York, NY, USA
2- The International Research Institute for Climate and Society, The Earth Institute, Columbia University, Palisades, NY, USA
3- Northern Medical University, Arkhangelsk, Russia
4-Joint Research Center, European Commission, Ispra, Varese, Italy

Date Created, Published or Revised: (Please specify which date is listed.) Overview / Citation / Dates

Amstislavski, P., Ceccato. P., Weedon, J., Chen, H., Zubov, L.
Surface Water Change and Access to Health Care Services in the Russian Arctic.
International Journal of Circumpolar Health. 2012 (in press).

Publication Information: (If publisher is different from the author, if a paper was published about the dataset use the journal name as the Publisher)

Publisher: Overview / Citation Contacts / Resource Citation Contacts (specify Publisher as Role)

Publisher location: Overview / Citation Contacts / Resource Citation Contacts

Series name (if any): Overview / Citation /Series

Issue (if Series is used): Overview / Citation / Series

Online Links [URLs] to supporting documents: Overview / Citation Contacts / Contact (add New Online Resource)

b) What geographic area does the data set cover?

Nenets Autonomous Okrug (Russian: Не́нецкий автоно́мный о́круг), Russian Federation.

Bounding box coordinates: (What is the bounding box of the dataset? Please use negative numbers for degrees West longitude and for degrees South, positive numbers for degrees East longitude and for degrees North latitude) Overview / Item Description / Bounding Box. How to use "Extent contains the resource"

68°50′N 54°50′E

Location Keyword: Overview / Topics & Keywords / Place Keywords

Nenets Autonomous Okrug, Не́нецкий автоно́мный о́круг), Russian Federation

c) What does it look like?

Maps of water bodies and ground.

Optionally, include any URLs to graphics or supply example graphics with this metadata record. Overview / Item Description / Thumbnail (for embedded graphics); Resource / Details / New Browse Graphic (for external graphics)

d) Does the data set describe conditions during a particular time period?

Time Period of Content: (Generally a start and end date/time of the dataset). Please use the UTC timezone. List what time zone the data is recorded in and the UTC offset. Resource / Extents / Temporal Period Extent. Specify time zone and UTC offset in Resource / Details / Description? Not sure how to indicate time zone in ArcGIS metadata editor. May need to add this after the XML is created by ArcGIS.

Annual maps of permanent water bodies detection for the period January 1 2004 – December 31 2009.

Should we indicate temporal precision of data (e.g., 1 day)?

Yes, year foe each map.

Currentness Reference: (Ex: The dates/times of the dataset refer to CTD casts, bottom grab times, bird observation time, referred time/date being described in Local Traditional Knowledge testimony) Resource / Details / Description?

e) What is the general form of this data set?

Shapefiles produced from derived from .HDF files of MODIS scenes covering the region over the period 2004—2009.

Data Presentation Form: digital, paper, samples. (In most cases, the answer is digital form) Overview / Citation / Presentation Form. If we're submitting shapefiles, for Presentation Form shall we use "Digital Map", and for FGDC Form shall we use "Vector Digital Data"?

Digital.

f) How does the data set represent geographic features?

Direct Spatial Reference Method: (Single word description of spatial references in the dataset. Can be one of Point, Vector or Raster. Ex: XBT/CTD casts may be considered Point locations, cruise track lines are composed of connected points (to form Lines/Vectors) and Raster is typically used for gridded datasets (model output, satellite, etc). If the dataset does not strictly conform to any of these types, provide an Indirect Spatial Reference instead.) Resource / Spatial Reference Representation / Vector Representation (Topology Level: Geometry Only; Geometric Objects: {Feature Count}; Point | Curve | Solid?)
Vector, polygon.

Indirect Spatial Reference: (Use this area to describe how to spatially define this dataset if it cannot be defined by a typical method above) Resource / Spatial Reference Representation / Indirect Representation

g) What coordinate system is used to represent geographic features?

Projection information: Specify the projection of the data, including the projection parameters specific to the named projection. Resource / Spatial Reference / Reference System (accepts codes, but not spatial ref details)

WGS 1984 UTM Zone 38N.

h) How does the data set describe geographic features?

How are the features defined? Is it stored in a particular format? (e.g., CSV, Excel, database) Don't know where to specify this.

Shapefile.

Within these data or file formats, how is the information defined? What are the columns or variable names used? What are the units for each column?

For each column or variable, describe: Fields / Entity and Attribute Details / Attribute

Name: variable or column name Label

GRIDCODE=1 water detected, GRIDCODE=2 ground detected.

Units: if applicable Definition? Unrepresentable Domain?

Meters

Valid Range: valid range of values or enumerated list Enumerated Domain | Range Domain

Missing value: if a value is missing, how is that shown in the dataset? Definition?

2) Who produced the data set?
   a) Who are the originators of the data set?

Author: (repeated above, may skip, included for completeness)

   b) Who also contributed to the data set?

Data Set Credit: Recognition of those who contributed to the data set. Overview / Citation Contacts / Resource Citation Contacts

Philippe Amstislavski[1]
Leonid Zubov[3]
Herman Chen, MBA[1]
Pietro Ceccato[2]
Jean-Francois Pekel[4]
Jeremy Weedon[1]

   c) To whom should users address questions about the data?

Point of Contact: for this dataset **Name, Organization, Phone, Email, etc**. Resource / Points of Contact / Contacts

Philippe Amstislavski, PhDAssistant Professor in the Department of Environmental and Occupational Health Sciences, School of Public Health
tel: (718) 270-2105 • fax: (718) 270-7285 • e-mail: amstislavski@gmail.com

3) Why was the data set created?

To identify water surfaces from remotely sensed data and demonstrate the integration of the remotely sensed imagery data acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors with data on human activity at the local communities level to contribute to the understanding the coastal ecosystems of the Arctic Ocean.

Purpose: (repeated from above, may skip) Why were the data collected?

4) How was the data set created?
   a) What methods were used to collect the data?

Methodology: Fields / Entity and Attribute Details / Attribute / Definition?

   b) From what previous works were the data drawn?

Source Information:

   c) How were the data generated, processed, and modified?

Process Steps: Fields / Entity and Attribute Details / Attribute / Definition?

The main challenge to identify water surfaces from remotely sensed data is the high variability of their spectral signatures. The spectral property of water is determined by the electromagnetic interaction of light with the constituent components of water via absorption or scattering processes either within the water column or at the water surface or on the bottom of the water. Theses constituents are: the phytoplankton (chlorophyll-a), the suspended sediments (i.e. solid particulate matter) resulting from erosion process and the colored dissolved organic matter (CDOM) resulting from the degradation of biological organisms. All these constituents vary in character and amount with the limnological/optical types, season, cyclic change of biological activity and human impact, and play an important role in determining intensity of the absorption and scattering processes (Xiaoling Chen and Zhifeng Yu, 2009). Consequently, the water-leaving radiance detected by the sensor shows great spatial and temporal variability, which makes its reliable discrimination particularly difficult (Gond et al. 2004).
In the framework of this study, a methodology recently proposed by Pekel et al. (2012) to detect open water surfaces on near real time was used. The method is based on a colorimetric approach of the signal. The rationale is that the perceived color in a color composition is directly determined by the shape of the spectral signature of the target. Consequently, it is possible to associate a land surface type to a range of colors.

The method includes three steps. (i) First, the daily MODIS surface reflectance images were composited on a 10-day basis (1 to 10, 11 to 20 and 21 to the end of each month) using the mean compositing algorithm (Vancutsem et al. 2007a; 2007b). This procedure averages the cloud free reflectance values from both Aqua and Terra after a quality control based on the MODIS standard data quality flags, and allows producing consistent (both spatially and temporally) cloud free imagery. (ii) Second, the 10-day MIR, NIR, and Red bands were assigned respectively to the Red, Green, and Blue components of the RGB color space and transformed into the Hue, Saturation, and Value components of the HSV color space based on a standardized colorimetric transformation as defined by Smith (1978). (iii) Finally, thanks to a large sampling of "open water surfaces" and "other land surfaces" spread both in time (i.e. different seasons and different years) and space (i.e. different land surfaces types), the spectral signatures and the colorimetric properties of these two clusters were characterized. Therefore a set of thresholds were identified (see Pekel et al, 2012 for more details) and applied on the Hue and Value subspaces to detect the open water surfaces on a 10-day basis at 250m (Fig. 1).

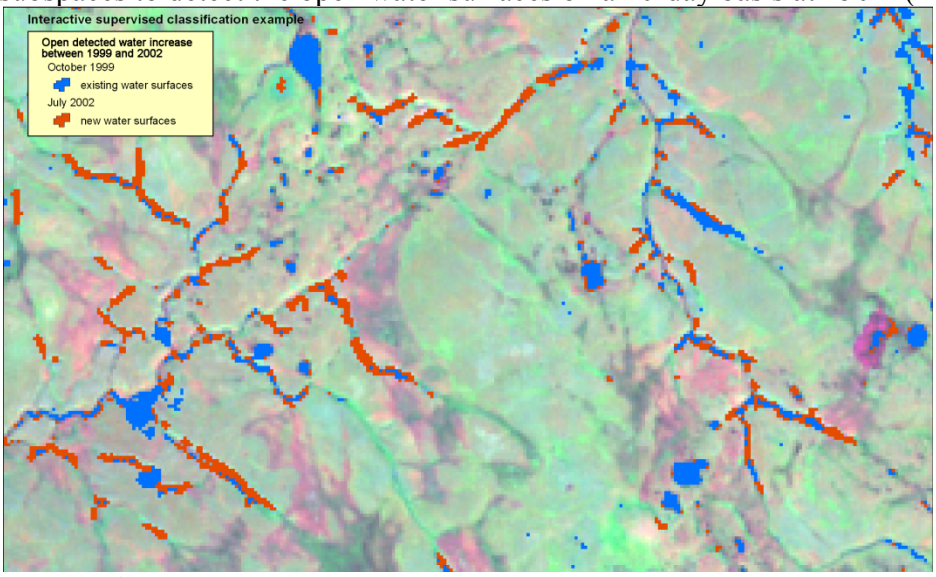

Figure 1: Open detected water increase between 1999 and 2002 at a study site in Kanin tundra of northern Russia using MODIS data product.

However, false water surface detections, which manifest by a low temporal frequency, were observed. These false detections resulted from the remaining perturbations after atmospheric correction and cloud removal. Indeed, over our area of interest the MODIS standard flags were not masking all the contaminated values that were then included in the compositing process and induced reflectance instability and consequently some false detection afterwards. In order to discard these false detections from our analysis, we considered only the water surfaces showing an occurrency (i.e. number of water detections divided by the number of available observation) above 35% on an annual basis. Indeed, as the false detections occurred randomly, both spatially and temporally with a low temporal frequency, the lower occurrency concentrated the false detections.

d) What similar or related data should the user be aware of?

Optionally, indicate information about other, related data sets that are likely to be of interest; or papers using this data set. Overview / Citation / Other Details, or Resource / References / Aggregate Information (if this work is part of a larger work)

5) How reliable are the data; what problems remain in the data set?

One drawback of this approach is that temporary water surfaces also characterized by low annual occurrence (e.g., water bodies that occur in less than 35% of MODIS scenes) were discarded. It decreased artificially the number of water surfaces considered in our study, but as the same criterion were applied for each year, and as we used theses detections in a relative way, it should not impact significantly on our results.

a) How well have the observations been checked?

Data Accuracy Report: (FGDC calls this Attribute Accuracy Report.) Explanation of the accuracy of the identification of columns/variables and assignment of values in the data set. The report can reference more extensive descriptions in other documents. (e.g. Calibration sheets, etc. Are there accuracy considerations that need to be stated about any of the particular data columns that might show up in subsequent quality control checks that might be explained by instrument or equipment changes or quirks?) Resource / Fields / Attribute / Value Accuracy

b) How accurate are the geographic locations? Resource / Extents / Description?

Horizontal Positional Accuracy Report: Explanation of the accuracy of the horizontal coordinate measurements (e.g. GPS location) and a description of the tests used.

250 meter.

c) How accurate are the heights or depths? Resource / Extents / Description?

Vertical Positional Accuracy Report: Explanation of the accuracy of the vertical coordinate measurements and a description of the tests used.

d) Where are the gaps in the data? What is missing?

No.

Logical Consistency Report: Explanation of the fidelity of relationships in the dataset and tests used. (e.g. What gross quality control checks were done to ensure data integrity? Example: Data and/or locations seem correct upon visual inspection when plotted on graphs/maps. A large data gap exists due to the elimination of data based on such and such a test). The report can reference more extensive descriptions

in other documents. If creating a Logical Consistency Report is not logical, use the default "not applicable". Should be under Resource / Quality / Report, but Logical Consistency Report is not an option for record type.

e) How consistent are the relationships among the data, including topology?

Completeness Report: Information about omissions, selection criteria, generalizations, definitions used, and other rules used to derive the data set. The report can reference more extensive descriptions in other documents. Should be under Resource / Quality / Report, but Completeness Report is not an option for record type.

6) How can someone get a copy of the data set?
   a) Are there legal restrictions on access or use of the data? What would be the typical constraints for our data?

      Contact Philippe Amstislavski

   Access Constraints: Restrictions on accessing or obtaining the data set. Resource / Constraints / Legal Constraints / Access Constraints

   Use Constraints: Restrictions on using the data set after access is granted. Resource / Constraints / Legal Constraints / Use Constraints

   b) Who distributes the data? Resource / Distribution

   This is the contact information for the dataset, if different from the Author, should data management get a question about the dataset we cannot answer.

   c) What's the catalog number I need to order this data set? Resource / Distribution / Distributor / Ordering Process / Ordering Instructions

   Optional, if the dataset has a unique identifier.

   d) What legal disclaimers am I supposed to read?

   Distribution Liability: What legal disclaimers are the user supposed to read? Resource / Constraints / Legal Constraints / Other Constraints?

   e) How can I download or order the data? Resource / Distribution / Distributor / Ordering Process / Ordering Instructions

   If there is a way to download or order the data than having to use the contact information above, include that information here.

7) Who wrote the metadata?

Philippe Amstislavski

8)

Metadata Author: (If different from the Author of the dataset.) Metadata / Contacts / Contact (specify Metadata creator as Position, specify Processor as Role)?

9) What other elements are required?

Status: (Status of the dataset. Is it in a completed form or is more information expected later, are updates expected if errors are found or corrections made?) Resource / Maintenance / Update Frequency?

10) **Taxonomy (if applicable)**

To include taxonomy in metadata records using this form we request a text file with a listing of scientific names. **A separate file is needed for each kingdom**. To make things easier, please use the kingdom name as part of the filename.

Example: **Filename = Plantae.txt**

```
name
Acer saccharum
Acer nigrum
Acer barbatum
Nyssa sylvatica
Liriodendron tulipifera
Fraxinus americana
Juglans nigra
Quercus alba
Quercus coccinea
Quercus velutina
Quercus rubra
```

Using these files, we will use the ITIS Tools page to convert the above scientific names to SGML/XML that we import to the FGDC metadata record to list taxonomy. You may use this tool in advance to check and quality control your taxonomic information. Please consult these ITIS web pages:

ITIS Tools: http://www.itis.gov/taxmatch_ftp.html

ITIS Web Page: http://www.itis.gov/

If you have specific scientific disputes or questions with information in the ITIS database please take it up with them. We currently do not have the resources to affect changes to the ITIS database.