CEMA Traditional Ecological Knowledge Data Standard

Drafted by Trevor Wiens, Apropos Information Systems Inc.(2012-04-27)

Introduction

The Cumulative Environmental Management Association (CEMA) has gathered Traditional Ecological Knowledge or TEK as part of its studies for its working groups for Surface Water, Ground Water, Air, Land, and Reclamation. Because of the unique role TEK plays in all these areas, CEMA has a dedicated Traditional Knowledge Working Group responsible for providing assistance and guidance to other working groups in the effective and appropriate use of TEK.

The original TEK research guidelines for CEMA written in 2006 lacked specific data standards. Data standards are essential to ensure that data gathered will meet both immediate project needs and long term use needs. CEMA is currently in process in updating these guidelines and this document provides data specific details.

Starting in 2010, CEMA started working with Apropos to evaluate if the Land Occupancy and Use Information System or LOUIS (http://www.aproposinfosystems.com/products/louis) would be useful for CEMA to store and manage its TEK data. After carefully evaluating LOUIS at the proof of concept stage and then with a feasibility study, CEMA decided in 2011 to proceed with the implementation of LOUIS for CEMA. In 2011 all past studies were evaluated, cleaned and imported into LOUIS. In February and March of 2012, Apropos trained CEMA Program Administrators in the use and operation of LOUIS. CEMA now has live search access to all past TEK studies and the ability to quickly and easily generate reports and maps from that data without the need of desktop GIS tools or GIS technicians.

What is TEK?

TEK is one of the many names used to describe the complex knowledge and belief systems of indigenous peoples. For cultural researchers the term TEK has a very specific meaning, but for most people it is synonymous with other terms such as Traditional Knowledge, Indigenous Knowledge, Traditional Land Use, Use and Occupancy, or Local Knowledge. For simplicity, TEK will be used in this document in the informal sense meaning the information gathered by CEMA in its cultural research efforts.

Purpose of this Document

This document is the authoritative document for CEMA TEK data standards. All aspects of this standard are based on industry best practises and are designed to ensure the highest possible quality data is gathered. Only high quality, well documented data are acceptable to CEMA because of CEMA's respect for community knowledge holders and CEMA's commitment to gathering data that are useful for both immediate and long term needs.

The details provided in this document are specific and all contractors to CEMA are required to follow them. Deviation from these standards without prior written authorization from CEMA will be considered a failure to meet contract requirements.

Data Management Overview

The core operating principle for proper data management is effective and accurate documentation. This entails documentation about how data are planned to be gathered, how the data actually were gathered, and definitions and use guidelines for the resulting data. Sadly most organizations fail in all three areas because success is only possible if it starts with well documented planning. After the fact documentation is seldom accurate and usually indicates insufficient planning. Therefore central to the new data standard for CEMA TEK projects is detailed methods documentation being presented to CEMA before initiating data collection. This should be followed up by a report on field deviations from the proposed methods with detailed accounting for why deviations occurred and possible implications on final results as a result of those deviations. The final report should contain all previous steps along with documentation, as appropriate in the final analysis of how variance in field methods may or may not have impacted the final results.

Data Quality

It is critical to understand that cultural research data quality is set and cannot be improved after the completion of the interview. Similarly it is also important to remember that TEK data are heavily context dependent. Given these two indisputable facts, it is imperative that the format of the data capture mimic the source as closely as possible to preserve the details and context in which they are given.

It is unfortunate that all too often complex and nuanced conversations about use in one location are reduced to a simple point and a one word code. This kind of simplification is disrespectful of the knowledge holder and replaces the original data with an oversimplified interpretation. LOUIS is designed to allow retention of context and relevant details so that data can be analyzed in terms of simple coding, complex searches, or within the full context of a conversation.

The effort to ensure quality data must start with careful planning and detailed documentation before talking to a single participant. The CEMA data standard and CEMA TEK database, LOUIS, are based on general good data management practises and on many of the cultural research concepts found in the book *Living Proof* (Tobias, 2009). This book focuses on Use and Occupancy mapping which is the primary type of research that CEMA has done in the past. *Living Proof* is an excellent guide for planning and conducting high quality Use and Occupancy Mapping research. The concepts developed for this discipline spill over nicely for other aspects of cultural research and thus form a solid foundation for CEMA's TEK data standards.

Consent and Ownership

Informed Consent

Following best practises and ethics in other social sciences, it is very important to make sure that participants understand that all TEK data gathered in projects funded by CEMA will be returned to and held by CEMA. The contractors must clearly understand that as contractors to CEMA they are acting as agents of CEMA so any information shared with them is shared with CEMA. Contractors must also understand that they are not to retain any TEK data after the completion of the contract. Contractors must clearly communicate this message to participants so that project participants fully understand with whom they are sharing the information under the terms of the TEK data sharing agreement.

Prior to 2012, CEMA had no means to effectively re-use the source data from TEK reports. Now that CEMA is a LOUIS user, all future TEK projects will be imported into LOUIS either after completion

or possibly during the data collection process. Again during the consent process, contractors must be clear with participants that data will be used in dual manner: for the current report and imported into LOUIS for ongoing use within CEMA as per the terms of their TEK data sharing agreements.

Apropos strongly recommends that CEMA adopt a standard consent form and process that names the contractor as an agent of CEMA. Ideally this consent process should be done by the TK Coordinator to ensure a consistent message is relayed to community members.

One approach to help keep the documentation simpler for participants is separating the TEK use agreement from the consent form so that each document can be kept short and clear. Some examples of this can be found in *Living Proof* (Tobias, 2009).

Data Ownership

Data shared with CEMA belongs to the knowledge holders who provided it under the terms of the TEK data sharing agreement, but CEMA has the right and responsibility to house and use those data. Contractors must ensure that they understand and operate in such a fashion that all information is transferred to CEMA for archiving and management. Contractors must also understand that they have no legal right to the information after the completion of the project and should put in place measures to remove all TEK from their systems after the completion of the project.

Data Gathering Methods

Methods Documentation and Data Review

Before data is gathered, contractors should have their data collection methods reviewed together by CEMA's TK Coordinator and CEMA's LOUIS data manager to ensure that their methods meet academic and ethical standards as well as CEMA's data standards and needs. In Chapter 15 of *Living Proof* (Tobias, 2009) an example of a Data Collection Manual for a research project is provided. Contractors are strongly encouraged to use this style of methods documentation and not deviate without good cause and written approval from CEMA.

Upon completion of the data gathering phase and before report production, the data should be reviewed by CEMA to confirm that agreed upon methods were followed and no significant errors in the data exist that will skew results or make data difficult to import and use in LOUIS.

Audio Recording and Transcripts

All interviews should be recorded using a high quality data recorder so that the full conversation is audible. Full transcripts may not be necessary but partial or non-verbatim transcripts are required. LOUIS has a powerful classification and search engine so even partial or non-verbatim transcripts can be used effectively.

Audio can be recorded using the tools of choice of the research team. However audio files are to be converted to .wav format and stored on CDs or DVDs for archival purposes when delivered to CEMA. In chapter 11 of *Living Proof* (Terry Tobias, 2009) there are a variety of practical recommendations on how to gather audio recordings that merit consideration.

At the top of each transcript a descriptive header is required. Required information is the CEMA contract number and project name, the date, start and end times of the interview, the location of the interview, the full name of the interviewer, the full name of the translator if one is used, and the full name of the participants. Participant community affiliations at the time of the interview should also be included along with any other information pertinent to the interview in general. This information ensures that the interview file will remain valid and useful if it gets renamed or moved to an

unexpected location within the CEMA file system.

Format examples for transcripts and attribute inclusion to ensure compatibility with LOUIS are included in Appendix A.

Date Information

No other single data field is cause for more misunderstanding than date information. LOUIS only supports the use of the ISO standard date format which is YYYY-MM-DD. It is required for CEMA TEK projects use of either the ISO format or the a variant of the ISO format using a three letter month as in YYYY-MMM-DD. Examples of these would be 2012-04-23 or 2012-APR-23.

Period of use

A common source of misunderstandings between scientific data and TEK relates to the issue of time. It is not uncommon for these disagreements to be resolved by understanding the reference time frames used by the different parties. To address this LOUIS supports user periods in terms of years, and / or months of the year. When developing the research plan, researchers need to consider the difficulty associated with temporal recall and plan accordingly. There is an excellent discussion of this issue in Chapter 9 of *Living Proof* (Tobias, 2009) that merits review and consideration at the design stage of any study. The short version of that discussion is that temporal recall is difficult and should be kept simple if possible. Often time periods relating to significant community events like “before and after the railroad” are useful and reliable. Another common practise is to record “in living memory”. Whatever temporal parameters are chosen, they must suit the purpose of the study and be explicitly stated in the interview and defined in the study methods documentation.

Clear Links between Maps and Text

During the interview process when maps are involved it is important to note verbally on the audio recording and mark clearly on the map using clear unique codes so that links can be made between the verbal discussion (and transcript) and the features on the map. If this is not done, all the contextual information provided by the interview participants is lost and the rich data set is reduced to simple interpreted codes. It is important to capture the original information as closely and as accurately as possible. Chapter 12 of *Living Proof* (Tobias, 2009) provides an excellent coding system for marking maps that should ensure both spatial and contextual accuracy. Contractors not willing to follow this system are required to get approval from CEMA in writing prior to starting data collection after demonstrating to CEMA that their preferred method will produce similarly high quality results.

A summary of this coding system is included as Appendix B. Please refer to *Living Proof* for additional details.

Clear Identification of Interviewees

In individual interviews it is easy to identify the speaker but in group sessions keeping these connections clear can be difficult. Contractors are strongly encouraged to put forth effort in group interviews to clearly identify on the audio recording who is or was speaking.

Clear Links between Photos / Video and Places and People

Any photos about places or people need to be clearly identified. Use of the same codes used to identify map features is required so that photos can be loaded with their associated map features and text. When doing GPS field work, photos should be taken of all sites as well as representative photos of trails and areas. These photos should all be archived in .tiff format on CD or DVD and a spreadsheet or csv file

with a listing identifying the photo file name, description, and map id is required.

GIS Data

GIS data files not compliant with CEMA standards or with missing or incomplete documentation will be considered undelivered.

Data must have analytical utility and meet CEMA standards

Arrows on a map cannot be effectively analyzed in a GIS environment. Therefore in cases of data such as wildlife movements, delineation of areas by season instead of generalized arrows has much greater analytical utility. Contractors are reminded to carefully consider long term and analytical implications in all data collection decisions.

GIS Standards

Like other aspects of computer technology, with GIS there are real certified standards and common local practises which are often mistaken for standards. Use of only documented and widely supported standards is the best approach to ensure usability of data on an ongoing basis. The following formats only will be accepted:

- Vector: ESRI Shapefile or GML with one layer per file
- Raster: GeoTIFF

For large scale maps it is required to use the appropriate UTM projection, usually zone 12, with the NAD83 datum using EPSG defined standards. For small scale maps use of the Alberta 10TM NAD83 projection is required. The respective EPSG codes are UTM Zone 11 - EPSG:2955, UTM Zone 12 - EPSG:2956, and Alberta Wide 10TM – EPGS:3400. Only these projections will be accepted

GIS Feature Attributes

To ensure that projection data is handled correctly LOUIS only imports spatial data from KML files. The conversion from ESRI Shapefile or GML files to KML format will be done by the CEMA data manager. The CEMA contractor must ensure that each attribute, be it line, point, or area has an attribute called section_code. The section_code attribute must contain the map code to link the map feature to the interview transcript. It is advised to not have attributes in the GIS file that duplicate the contents of the text file (with the exception of section_code) to ensure that values in the text file are not over-written by values in the GIS file during import into LOUIS.

GIS Metadata

Many GIS software platforms will automatically produce standards compliant ISO 19115/19136 XML format. Unfortunately the default information in these files has little value. All GIS files submitted should be complemented with complete ISO 19115/19136 XML standard metadata files with full process step, attribute, contact, and data quality information.

GPS Data

A variety of qualities of hand-held GPS units are available. Use of better quality units or post-processing of data is recommended to improve the accuracy of the GPS data. In all cases the GPS unit model, PDOP for that day and location, and other details including original GPS files must be provided so the accuracy of the information can be adequately assessed. Free software is available from Trimble (http://www.trimble.com/planningsoftware_ts.asp) which can calculate PDOP values after the fact or in advance for planning best times for GPS data collection.

Summary

Table1: *Core Documents*

Required Documents	**Time Required**	**Reference Examples**
Summary of Past Methods	In RFP	N/A
Data Collection Manual - Draft	Before starting data collection	Living Proof Chapter 15
Updated Data Collection Manual and data collection report	After data collection and before analysis	N/A
Final Report including previous documentation – Draft	Before final invoice	N/A
Final Report	After approval of draft by CEMA	N/A

Table 2: *General Standards*

Content	**CEMA Standard**
GIS Files	ESRI Shapefile, GML (on layer per file), GeoTIFF
Projections	ESPG:2955, 2956, or 3400
GIS metadata	ISO 19115/19136 XML format
Date Information	Preferred: YYYY-MM-DD eg: 2012-04-23 Accepted: Use of letter month YYYY-MMM-DD eg. 2012-APR-23

Appendix A – LOUIS Data Formats

The Land Occupancy and Use Information System or LOUIS allows users to enter data directly or import it. In many situations because of a certain amount of discomfort with technology by many knowledge holders, the pen, paper map, and cassette tape recorder are the most effective tools for gathering information. In this case importing data becomes an important aspect of planning a research project.

In addition to support for spatial data and transcripts, support for photographs for sites and individuals is included LOUIS version 1.3 to released in mid-May 2012. Audio support will be added in version 1.4 which is scheduled for release in July of 2012. CEMA contractors should consider how the ability to link audio recordings to the interview transcripts in LOUIS and the ability to add photographs to specific locations are of long term use and plan accordingly.

The interview transcript will likely be recorded and stored in a word processing document and will likely be similar to the material in Sample Listing 1.

Sample Listing 1

```
I: I would like to ask you about where you hunt moose?
P: OK. Well... my favourite spot is here.
I: Note: I'm marking this location as MS001. Why do you like this spot?
P: I'm not so young any more so I like to take my boat so I don't have to carry
the moose too far. This spot is close to the lake and it has lots of moose.
I: Thanks. Are there other places you like to hunt moose?
P: When the lake is rough I take my truck and hunt here instead.
I: Note: I'm marking this location as MS002
```

You will note that the codes used unlike those recommended in *Living Proof* (Tobias, 2009) have leading zeros. Use of leading zeros is recommended by Apropos because it makes certain aspects of computerized data handling more straightforward but it is not strictly required.

In LOUIS this single flow of text is broken into parts based on the unique spatial feature discussed or in the case of non-spatial data, the topical content of the conversation. Breaking the conversation into chunks is a common part of an interview coding process and allows the LOUIS search engine to be more specific and accurate. LOUIS also allows researchers to add additional attributes such as security codes, map source and scale and other information. The text from Sample Listing 1 is shown in Sample Listing 2 after being transformed to include tags for LOUIS.

Sample Listing 2

```
[section]
section_code=MS001
security_code=PR
section_text=I: I would like to ask you about where you hunt moose?
    P: OK. Well... my favourite spot is here.
    I: Note: I'm marking this location as MS001. Why do you like this spot?
    P: I'm not so young any more so I like to take my boat so I don't have to
      carry the moose too far. This spot is close to the lake and it has lots of
      moose.
spatial_source=PM
spatial_scale=50000

[section]
section_code=MS002
security_code=PR
```

```
section_text=I: Thanks. Are there other places you like to hunt moose?
    P: When the lake is rough I take my truck and hunt here instead.
    I: Note: I'm marking this location as MS002
spatial_source=PM
spatial_scale=50000
```

In this case the sections have been marked PR for Private. This means only high trusted staff along with the data entry staff will have permission to see this text and its associated attributes including the spatial features referenced. The other information is that the data was gathered on a paper map (PM) of a 1:50,000 scale. If codes are not included in the file LOUIS will assign the default values for that field. These field values can also be added to the spatial file if that is easier for the research team.

Contractors are not required to transform their transcripts to this format, but they are encouraged to consider that this will be done for 3 reasons:

1. By transforming the data it can be easily imported into LOUIS and if contractor staff are trained in the use of LOUIS, LOUIS' automated coding and mapping features can simplify the process of the final report.
2. By using these codes, research teams are reminded to keep track of these different aspects of the data they are collecting and record them in a timely and accurate fashion.
3. If the data is close to import ready, it is easier and quicker for CEMA to evaluate research data and ensure compliance with CEMA's standards.

LOUIS attributes and valid values are listed in Table 3.

***Table 3**: LOUIS codes and valid values. Defaults are in bold*

Content Attribute	Status	Valid Values
section_code	Required	Any alphanumeric code up to 10 characters in length
security_code	Strongly recommended	PU – public CO – community **RS – restricted** PR – private
section_text	Required	Transcript text
measure_value	Optional	Integers greater than zero or float values depending on the measure_type
measure_type	Optional	**P – presence** O – ordinal C – count F – continuous

Table 3 continued

Content Attribute	Status	Valid Values
use_period	Optional	R – refused U – unknown **N – not recorded** P – provided
use_period_start	Optional	ISO formatted date representing the start at which this feature or area started being used
use_period_end	Optional	ISO formatted date representing the start at which this feature or area stopped being used
annual_variation	Optional	R – refused U – unknown **N – not recorded** SP – sporadic SE – seasonal Y – all year
annual_variation_months	Optional	A comma separated list of month numbers indicating the months of use
spatial_source	Recommended	CG – corrected GPS HG – hand-held GPS **OS – on screen** PM – paper map
spatial_scale	Required if PM used for spatial_source, otherwise ignored	Integer greater than zero for the denominator in a maps representative fraction.
note	Optional	Textual comments
sequence_number	Optional	A unique integer representing the sequential order of an interviews sections
tags	Optional	A comma separated list of user defined tags to describe the content.

Appendix B – Map Coding Methods in Brief

In TEK research the common and usually confusing standard method for marking maps has involved multiple colours and codes to make marks on the maps with little to no regard for connecting those mapped features to the interview transcript. This approach is not acceptable because it replaces the participants' source information with the interviewer's interpretation.

In addition to replacing the original data with interpretations, the use of a complex coding book with a dozen coloured markers slows the process and is error prone.

A simpler and cleaner approach is possible that ensures that the interview transcript or recording can be linked to the map features for use and also to double check the accuracy of the code at any time in the future.

In Chapter 12 of *Living Proof*, Terry Tobias details an alphanumeric system which consists of 2 parts, first a two letter category code and second a number indicating the sequence of features marked. This appendix provides a very brief summary of the materials presented in *Living Proof* and the full methods detailed in *Living Proof* should be followed. Although for use in LOUIS, the category code is not strictly needed, its inclusion can be helpful for the interviewer and interviewee alike and is recommended.

The other aspect of the Tobias map marking method is using colours to discriminate between overlapping features on an as needed basis rather than trying to encode content with colour. This clarity focused approach to using colour reduces the likelihood of digitizing error after the data has been gathered.

Examples of the map codes are:

- MS101 – Moose and the 101st feature recorded during the interview
- XB23 – Other (eXtra) Bird and the 23rd feature recorded during the interview
- CA1 – Cabin and the 1st feature recorded during the interview

The paper map would then have a line, point, or area marked on the map and then a leader line from the feature to the code to make it clear. Underlining all codes of lines features as suggested in *Living Proof* can improve clarity when digitizing the paper maps. See Table 17 *Living Proof* in for a good visual summary.

Other considerations

- Be aware of the pens used to mark maps. Use of fine tipped pens of indelible ink (0.6mm) is recommended because this improves the precision with which paper maps are marked and greatly reduces the possibility of smudging.
- Although not recommended in *Living Proof*, putting leading zeros in front of the sequential number can improve the clarity of the codes when they are later handled by GIS and database staff. Examples of the above with leading zeros would be MS101, XB023, CA001. Use of this number of leading zeros is based on the assumption that you will not mark more than 999 features in a single interview. If leading zeros are used it is ESSENTIAL that the alphanumeric codes are always 2 letters long so that there will be no confusion between a capital letter O and a zero. CEMA does not require the use of leading zeros and suggests that the data collection team consult with their GIS and data processing staff to determine their preference before submitting the draft Data Collection Manual to CEMA.

References

Tobias, Terry N., 2009. Living Proof: The essential data-collection guide for Indigenous Use-and-Occupancy map surveys. Ecotrust Canada and Union of BC Indian Chiefs, British Columbia, Canada. (http://ecotrust.ca/first-nations/new-book-use-and-occupancy-map-surveys-now-available)