Appendix H (Informative): CMECS Best Practices in Crosswalking

Introduction

One of the critical objectives of the Coastal and Marine Ecological Classification Standard (CMECS) is interoperability with other classifications and datasets. Once CMECS is endorsed as a Federal Geographic Data Committee (FGDC) standard, users receiving federal funds for classification of coastal and marine ecological units will be required to, at a minimum, crosswalk their source classification units to the CMECS standard. This guidance outlines the concepts and issues associated with crosswalking data from individual source classifications to the CMECS classification, and it also provides guidance on best practices for reporting source units into the CMECS framework.

Concepts

Crosswalking can be done to compare the conceptual units from one classification to units in another classification (unit-to-unit crosswalking) or relate observational data (samples, plots, or mapped polygons) that are collected using one classification to another classification (observation-to-unit crosswalking).

Unit-to-Unit Crosswalking

Unit-to-unit crosswalking is needed for identifying the relationship between individual units from a source classification and CMECS units. Unit-to-unit crosswalking requires a comparison of the concepts and circumscriptions of the units in both classifications in order to make the relationships between them explicit. Examples of situations where unit-to-unit crosswalking may be required include identifying the relationships between the units in a specific classification (the FGDC Wetland Classification [FGDC 1996] to CMECS) or e.g., assigning a CMECS unit name to a unit described in the literature that uses a different classification scheme.

Observation-to-Unit Crosswalking

Making the best use of existing data is a critical need for most users. Users have two potential pathways to take advantage of existing observation data for reporting to CMECS. These include re-analysis of the source data and re-interpretation of observation data.

Re-analysis of Source Observation Data

Re-analysis of source data requires *de novo* classification of the data using the CMECS framework; it does not require any crosswalking. Whenever feasible, re-analysis of the source data using the CMECS framework will result in the most accurate reporting of CMECS units, assuming the data needed to assign CMECS units are available in the source datasets.

Re-interpretation of Observation Data or Derived-Data Products

Re-interpretation of observation data or derived-data products (like maps) requires the user to make explicit the relationship between the observations (e.g., samples, plots, and polygons) and CMECS units. Perhaps the most common examples of re-interpretation of data products are (a)

crosswalking individual map polygons to CMECS or (b) labeling grab samples with CMECS units. Because observation data have specific measurable parameters associated with them (rather than a range for a conceptual unit description), they can often be more precisely attributed to a CMECS unit—provided the necessary data have been collected.

Crosswalking Methodology

Making the relationships between the units in the source classification and CMECS explicit is essential to effectively crosswalking the units. Comparison of units from one classification to another requires a thorough understanding of all the units, their thresholds, and any implementation conventions.

The following guidance suggests the minimum required metadata for the crosswalking process. This approach is an adaptation of the Taxonomic Data Working Group (Franz, Peet, and Weakley 2008) crosswalking model. The approach applies both to unit-to-unit crosswalking and to observation-to-unit crosswalking.

General Approach

- Wherever possible, review the original source data for the project. If these are available, they will allow more informed crosswalking from the source unit to the CMECS unit.
- Create a unit-to-unit crosswalk as a first step, employing the crosswalking approach described below to make the relationships explicit. Spreadsheets or tables are useful tools to help visualize the relationships. If the units are numerous and the relationships complex, creating a database that tracks the relationships between units may be a worthwhile effort.
- Automate where possible to help make the connections between observations and units. Develop algorithms for observation-to-unit crosswalks to facilitate crosswalking. However, it should be noted that some relationships can only be rectified with human interpretation of the source data.
- Incorporate other ancillary data if that meets your objectives. If source data for a study were not collected with CMECS thresholds in mind, identify other datasets that can be combined with the source data to help make the relationship to CMECS units clearer.
- When crosswalking observations to CMECS units, maintain any specific data that are at a finer scale than CMECS units and consider applying CMECS modifiers so the necessary detail is not lost.
- Only crosswalk to CMECS based on the source data you have. If the data are not available, it may be impossible to crosswalk some data.

Crosswalking Standard

Complete the following fields of information for each unit in the source classification. See Table H1 for an example of the standard applied to a unit-to-unit crosswalk between FGDC-STD-004 (FGDC 1996) classification substrate types and CMECS substrate units.

- **Unit Name in the Source Study** – Names of the units that are noted in the study using the source terminology.

- **Concept Reference for Source Unit** – Name and reference for the source classification.

- **CMECS Name** – The most closely related CMECS name(s) for the unit(s) identified in the study, from any level or any component that is related to the units in the first field. Users can list one-to-many CMECS names that relate to the data, preferably at the most appropriate level of the hierarchy.

- **Relationship of the Source Units to CMECS Unit(s)** – Description of the relationship between the unit(s) in the study and the CMECS unit name(s) listed with the following qualifiers:

    - *Equal (=)*: There is a one-to-one relationship between the source unit and the CMECS unit. The names of the units may or may not be the same. Units with "equal" relationships—but different names—are considered synonyms.

      Example: The terms "gravel" and "pebbles" are used in different classifications to mean the same thing.

    - *Nearly Equal* (≈): The source unit is nearly equivalent to the CMECS unit. The thresholds or concepts vary by a small—but insignificant—way for most practical applications.

      Example: The definition for "Mud" in the FGDC (1996) classification has a very small difference in the threshold for particle size than the "Undifferentiated Silt and Clay (mud)" unit in CMECS. The difference is less than 0.012 millimeters.

    - *Greater Than (>)*: The source unit is more broadly defined than the CMECS unit. The threshold range for the source unit definition is wider than the threshold range of the CMECS unit definition. The concept of the source unit fully contains the concept of the CMECS unit, but the source unit also includes additional entities.

      Example: There is a one-to-many relationship between the Florida SCHEME classification (Madley, Sargent, and Sargent 2002) for the "Reef/Hardbottom" class and the CMECS Coral Reef, Mollusk Reef, Worm Reef, and Bedrock units. The Reef/Hardbottom Class is greater than each of the related CMECS units.

    - *Less Than (<)*: The source unit is more finely classified than the CMECS unit. The threshold range for the source unit definition is narrower than the threshold

range of the CMECS unit definition. The source unit concept is fully included in the CMECS unit, but the latter concept contains additional entities.

Example: The FGDC (1996) "Bedrock" unit is finer than the CMECS "Bedrock" unit. The FGDC threshold for defining bedrock requires that the area be covered by at least 75% bedrock, whereas the CMECS definition requires the area to be covered by at least 50% bedrock. So although the names are the same, the concept definitions based on their threshold values are different.

- *Overlapping (><)*: The source unit is not clearly broader or finer than the CMECS unit. The two concepts contain at least one common entity, and each concept also contains at least one entity that the other does not contain. Neither concept is fully contained in the other.

- *No Equivalent (<>)*: The source unit does not have a clearly related unit in the CMECS classification.

  Example: The SCHEME (Madley, Sargent, and Sargent 2002) classification recognizes a type called "Platform Reef" that has no related concept in CMECS, and CMECS recognizes a "Microbial Mat" unit that has no related concept in SCHEME. This case most often arises when a source classification has a different objective than that of CMECS—or if the source classification has been developed to describe local, more detailed ecological units.

- *Unknown (?)*: The relationship between the source unit and CMECS units is unknown. Although some correlation is likely, a determination cannot be made from the data provided in the source study.

CMECS Relationship Confidence

For each value entered in the "Relationship to CMECS Unit" column, use the following values to indicate confidence in assigning the relationship between the CMECS unit names and the habitat units in the study.

- *Certain*: Either CMECS was used as the habitat classification for the study, or there is evidence (textual or data) in the study for the relationship between habitat names in the study and CMECS units. If exact CMECS names are not used, well-established synonyms or geographic associations make clear the relationship between CMECS units and the units in the study (e.g., gravel = pebbles).

- *Somewhat Certain*: The relationship between the habitat descriptions in the study and CMECS can be inferred, but is not explicitly stated and synonyms are not clear.

- *Not Certain*: There is a connection between the habitats in the study and CMECS habitats, but the relationship is not clear. Connection is based on the best educated guess by the party coding the study.

Relationship Notes

For each source unit, specify or clarify the relationships to related CMECS units. Identify where clear nesting of units occurs as well as any threshold differences.

Table H1 Example of Crosswalk of FGDC (1996) Substrate Classes and Subclasses to CMECS, applying the CMECS metadata standard.

Cowardin Class/Subclass	Relationship to CMECS	CMECS Class/ Subclass	Confidence	Relationship Notes
Rocky Shore	<	Rock Substrate	Certain	CMECS Rock substrate = Cowardin Rocky Shore + Rock Bottom. Shore is considered in the CMECS Geoform Component.
Bedrock	<	Bedrock	Certain	Thresholds vary.
Rubble	=	Boulder	Some-what certain	Cowardin Rubble = CMECS Boulder. (Boulder is under Unconsolidated Substrate in CMECS.) Cowardin does not specify size of boulders.
Unconsolidated Shore	<	Uncon-solidated Substrate	Certain	CMECS Unconsolidated Substrate = Cowardin Unconsolidated Shore +Unconsolidated Bottom. Shore is considered in the CMECS Geoform Component.

Table H2 (Continued) Example of Crosswalk of FGDC (1996) Substrate Classes and Subclasses to CMECS, applying the CMECS metadata standard.

Cowardin Class/Subclass	Relationship to CMECS	CMECS Class/ Subclass	Confidence	Relationship Notes
Cobble/Gravel	<	Coarse Uncon-solidated Substrates	Certain	Cowardin Cobble Gravel does not include Boulders.
Sand	≈	Sand	Certain	Very small (0.012 millimeters) lower threshold difference. Cowardin doesn't recognized Muddy Sand as a separate entity.
Mud	≈	Mud	Somewhat Certain	Cowardin doesn't recognize Sandy Mud as a separate entity.
Organic	≈	Organic Substrate	Somewhat Certain	
Rock Bottom	<	Rock Substrate	Certain	CMECS Rock Substrate = Cowardin Rocky Shore + Rock Bottom.
Unconsolidated Bottom	<	Uncon-solidated Substrate	Certain	CMECS Unconsolidated Substrate = Cowardin Unconsolidated Shore +Unconsolidated Bottom.

Map Crosswalking Challenges and Solutions

The list below outlines some of the specific complexities encountered when crosswalking a mapped observation to CMECS. In general, it is best to first label polygons with the most closely related CMECS unit(s), and then document the relationship and uncertainty. However, sometimes it will be necessary to report on units that are a level or two higher in the CMECS classification to find common ground.

- **Challenge: Units in the source map relate to more than one CMECS unit.** This occurs when the source unit concept is greater than the CMECS unit concept, but it also may occur when the project's minimum mapping unit is large (allowing for heterogeneous polygons).

  Solution: Report all of the CMECS units that are represented in the polygon.

- **Challenge: The source map recognizes more detailed polygons than CMECS.** Often local or regional map applications have specific objectives that require finer classification units than are described in CMECS. This is an example of the "less than" scenario described above.

  Solution: Label the polygon with the single related CMECS type. Also, consider labeling each polygon with a CMECS modifier (to avoid losing the detailed information in the map). If it is not necessary to maintain the detailed information, consider lumping together adjacent polygons that represent the same CMECS unit, where appropriate.

- **Challenge: The source map boundaries for a unit are located in a different place than the CMECS boundaries would be located.** Most often, this occurs when the source classification units and CMECS units vary in their thresholds for defining concepts. For example, CMECS defines the up-river boundary of an estuary as the head of tide, whereas FGDC (1996) defines it on the basis of salinity. In this case, the boundary line of the estuary on maps of these two classifications would not match. Here, the concept of "Estuarine" is greater in CMECS than in the FGDC (1996). On a map, this results in discordant boundaries.

  Solution 1: This essentially the same as the case where the source unit is either broader or finer (depending on the situation) than the CMECS unit. Keep the existing line and follow the protocols described in the scenarios above.

  Solution 2: If ancillary data that reflect the CMECS boundaries are available, incorporate them to modify the map to reflect the appropriate CMECS boundary.

- **Challenge: The relationships between the classification units are so complex that it is difficult to reassemble the map units into coherent CMECS classes.** Most often, this case occurs when the relationship between the source units and CMECS units are based on fundamentally different concepts—and the resulting many-to-many relationships make rectification extremely challenging.

Solution: If the relationships between the source unit and CMECS units are many-to-many, either apply the most closely related unit higher in the CMECS hierarchy and then indicate the complexity, or maintain the source units and document the intersections using the crosswalking methodology.

- **Challenge: Map units don't allow differentiation of CMECS units, because the source data does not reflect the necessary CMECS thresholds.** Datasets collected with another classification scheme in mind often do not include the information that is needed to identify CMECS units—or the scale of the source data aren't sufficiently resolved to identify CMECS units.

  Solution 1: Incorporate additional ancillary data that will allow attribution (within limits of project objectives).

  Solution 2: If ancillary data are not available and existing data do not support crosswalking to CMECS even at a higher level, do not proceed with the crosswalk. Instead, label the unit with the name from the source classification, and indicate that there is not enough information to determine a CMECS unit.

- **Challenge: The map unit has no apparent equivalent CMECS unit.**

  Solution: Check the CMECS modifiers to see if there is an equivalent. Check to see if there are units up a level or two in the CMECS classification to find common ground. If there is no related type higher in the hierarchy or in modifiers, label the unit with the name from the source classification, and indicate that there is no CMECS equivalent. Propose the new unit for inclusion into CMECS.

Literature Cited

FGDC (Federal Geographic Data Committee). 1996. FGDC-STD-004. *Classification of Wetlands and Deepwater Habitats of the United States*. Reston, VA. Federal Geographic Data Committee.

Franz, N., R. K. Peet, and A. S. Weakley. 2008. "On the Use of Taxonomic Concepts in Support of Biodiversity Research and Taxonomy." Section 5 in *The New Taxonomy*, 63–86. Systematics Association Special Volume 74. Symposium Proceedings. Edited by Q. D. Wheeler. Boca Raton, FL: Taylor & Francis. http://www.bio.unc.edu/faculty/peet/pubs/Cardiff.pdf.

Madley, K., A. B. Sargent, and F. J. Sargent. 2002. *Development of a System for Classification of Habitats in Estuarine and Marine Environments (SCHEME) for Florida*. Report to the U.S. Environmental Protection Agency Gulf of Mexico Program (Grant Assistance Agreement MX-97408100). St. Petersburg, FL: Florida Marine Research Institute, Florida Fish and Wildlife Conservation Commission.