

Reasoning for the Creation of a Unique ID in a Digital Bathymetry Database from Crowd-sourcing Systems

Submitted by Coordinator DFMCG

SUMMARY

Executive Summary: This document provides information on the reasoning for the creation of a unique ID in a digital bathymetry database and provides suggested formats to be use.

Action to be taken: 6

Related documents: None

1. In the effort of developing a Digital Bathymetry Database from Crowdsourcing Systems (i.e. SEA ID, Rose Point, Insight Genesis, Open Sea Map, others), a primary unique ID is the main key needed for all computer applications and by end users. With a unique ID the data can be managed and cannot be lost.
2. The first essential quality requirement IDs' need is to be unique; no two data entries in the crowdsourcing database should have the same ID. The IDs should be convenient to use, which means they should be, if possible, concise, distinctive, and sequential to uniquely identify each record in the database. In the need of having to tackle relational data structures, the unique IDs are a vital element in doing so. If the data need to be exported to another system or need to interact with other databases, having a key field that uniquely identifies each record in the database will guard against confusion or even possible loss of data integrity.
3. Any method that will produce a unique ID can be used. It is recommended that the key for the crowdsourcing bathymetry database should be preceded by a single text (i.e. 5 digit code) identifying the Crowdsourcing System, for example:
 - SEAID for the SEA ID Crowdsourcse Bathymetry Exchange Platform
 - ROSEP for the Coastal Explorer Rose Point ECS Crowdsourcing System
4. Following the alpha characters, a unique key to ID each vessel and data row entry should be created. From the point of database development, there are two methods to do that:

Method 1 (natural key): Generally, a key is composed or two or more data attributes that uniquely identifies a single record in a table. In a physical database a key would be

formed of one or more table columns whose value(s) uniquely identifies a row within a relational table. The key created from this method will mean something to users. Ideally, the key (i.e. alpha/numeric combination) should include a Vessel Identification Number, but this might not be feasible in all cases because of privacy concerns.

Method 2 (surrogate key): Using this method, keys don't have a natural relationship with the rest of the columns in a table. The surrogate key is just a value that is generated and then stored with the rest of the columns in a record. The key value is typically generated at run time right before the record is inserted into a table. One way to create this key is using the Universally Unique Identifier (UUID) generator, also known as GUIDs (Globally Unique Identifier) which is basically a 128 bit numbers presented in a hexadecimal grouped form that can guarantee uniqueness across space and time. GUIDs are statistically unique; the odds of two different clients generating the same GUID are infinitesimally small. The key created by this method will not have any meaning to the user, therefore has no risk of users changing the values of the primary keys, and the programming can be more consistent because all keys are of the same format.

5. Further discussion is needed to decide in the method to use to create a key. Nevertheless, in any method chosen, it is recommended that the unique ID (crowdsourcing system alpha character + key) should have:

- No leading spaces
- No embedded spaces
- UPPER CASE alpha characters to identify the crowdsourcing system
- If possible, no special characters (/, \, %, @, etc.).

6. The CSBWG is requested to note the information provided and take whatever action is deemed appropriate.