

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

Frequently Asked Questions (FAQs) about dbSEABED

What is dbSEABED?

- dbSEABED makes a single integrated data set from seabed sample information that has been collected over many years by ocean expeditions and research programs.
- dbSEABED is an “Information Processing” system that produces outputs that are compatible with Geographic Information System (GIS), Relational Database (RDB), as well as several other highly useful kinds of outputs. The outputs can be used in many end-user software applications.

What kind of outputs does dbSEABED produce?

- For mapping in Geographic Information Systems, dbSEABED produces a number of flat-files that are comma delimited. Those files give the 20 most useful parameters, and information on grain types and sedimentary features. These outputs can be enhanced in various ways, for instance by combining analytical and descriptive outputs, assigning gridded water depth values where not present, and re-arranging the data by its position within the seafloor.
- For use in Relational Database Systems like Oracle, MySQL or MS Access, dbSEABED produces a set of tables that are linked at the levels of seafloor site or sample using a system of numeric keys.

What quality control measures are in place for dbSEABED?

Quality control is implemented at many stages in dbSEABED:

- initially, by selecting which data sets to enter;
- by personnel scrutinizing the data at time of import, virtually item-by-item, for example for implausible values and positions not conforming to a report's survey pattern on a map; in some cases special filtering programs are written, for example to detect implausible ship speeds between stations;
- by testing in programs whether the data entries are text (string), numeric or a special formatted code as specified in the data model;
- in programs, testing if values lie within their plausible ranges, for example between -8 and 12 phi for grain sizes;
- by comparing site locations to land areas;

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

- in a more sophisticated way, comparing reported site water depths to a regional or global topography (such as etopo2);
- screening the program output values, also using plausibility filters;
- by having end-users of the data report any problems and having them fixed at source;

As each new data set is entered, it is tested in programs of dbSEABED to seek errors, etc. The program detections of problems are highlighted on screens during run time and are also reported to logs. As errors are detected, edits are made in the structured data files, complete with metadata explaining the edits. In some cases data will be deactivated (flagged out), again with metadata explaining.

What is the SRC file output?

- The “Source” (*SRC*) file gives basic information on the survey and data set that a sample belongs to, including the collector and institution, confidentiality, number of samples, dates, region, etc.

What is the function of the output CMP and FAC files?

- The *CMP* files provide information on the abundances (or prominence) of grain types, sedimentary features, and other components of the seafloor, for example of Halimeda algal grains, of ripples, and of hydrogen sulfide odor.
- The *FAC* file synthesizes these kinds of data by applying a Fuzzy Set operator and creating a superset across a selection of components/features – a geologic facies. For example, a calcareous pelagic facies is defined as the FST membership across samples that show nannoplankton, planktic foraminifera, or pteropod remains, or are described as calcareous ooze.

What is the CMP file output for?

- Geologists and others frequently describe the presence and abundance of components in a sediment, like shell debris, quartz, heavy minerals, etc. They do this by grain counts or in their descriptions. dbSEABED assesses the abundances of these components and outputs some of them in the *CMP* file. Users of the software can select which components will be output. Counting grains (and gauging their volumes or weights) is not an accurate process, and neither is conveying the abundance of a component in a description. The results in the *CMP* file are therefore only an approximate guide to abundances.
- The *CMP* file also carries fuzzy membership values for some features of the seabed, like odors (for example, H₂S), and bioturbation and ripples.

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

What is the meaning of the data in the FAC output file?

Once the Memberships (Abundances) of components are calculated, they can be combined into an assessment of the membership a sample shows to a range of sedimentary, rock or ecologic facies. Each Facies is typified by components A_f, B_f, \dots with memberships a_f, b_f, \dots , for example a CalcareousPelagic facies as "pfrm nan ptr cal_ooz ; 1 1 1 1 0 0". The components are senior synonyms from the parsing dictionary. Given a set of components in a sample A_s, B_s, \dots with memberships a_s, b_s, \dots , the sample membership of each facies is calculated as $\text{MIN}(a_c, a_f) + \text{MIN}(b_c, b_f) + \dots$ for each case of a coinciding component. In Fuzzy Set Theory this is a set intersection - an AND.

Of course, the Facies output is possible only from the parsed word-based data. Because not all studies report grain components, it is advisable to plot these results as point symbols only, not areal griddings.

What do all the “-99”'s mean?

- “-99” is the null value for numeric fields output by dbSEABED. Null values will be found only in the output tables.
- We have to specify null values because otherwise some end-user mapping and analysis applications make decisions about what a blank field in the table represents.
- The value “-99” is used because it almost never occurs in real data. It does for longitudes however, and for latitude/longitude the null is “-999”. The null for output string (text) data are “-“.

Why are there negative values for seabed strength and critical shear strength?

- These parameters are given in terms of their logarithm (base 10). A negative value implies a strength or shear stress less than 1.

Compared to values given in some scientific papers, the output values from dbSEABED are quite rounded. Why?

- Studies of uncertainty show that accuracies on most sediment parameters are of order 1-3% of the total parameter range even under favorable laboratory conditions. The significant figures in dbSEABED outputs reflect the observed uncertainties.

What is the Rock membership?

- Many observations of the seafloor report the presence of rock, loose and as bedrock. This output statistic is meant to convey the degree of exposure of rock which is coarser than cobble (-8 phi). Loose rocks and bedrock which are partly covered by sediment give memberships of less than 100%.

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

- Terms which have rock memberships include: basalt, granite, limestone, rock, boulders, lithified, hard, hard bottom, reef, rock platform, greywacke, where they occur in LTH or SFT themes, that is are not part of a grain-type description as in a PET theme line. For output data on specific rock types like basalt, refer to the CMP output file.
- Note that samples which have only loose sediments recorded are given a value of zero rock membership in CLC outputs.

What is the meaning of the SeaBedClass and ClassMembership?

These columns put out the facies showing the largest membership value for each sample, and that largest membership. The memberships are calculated as described in the FAQ about the FAC file. Output occurs only provided that the membership is greater than or equal to 0.33. Of course, this output is possible only from the parsed word-based data.

What is the Weed membership?

- Many observations of the seafloor report the presence of soft algae, seagrasses in general terms or specific to a taxon. Depending on the associated abundance terms like sparse, abundant, meadows, scattered, etc, the weed is given a membership which is reported to outputs (PRS). There are no EXT outputs for weed membership. Calcareous algae like Halimeda and Lithothamnion are not included.
- Terms which have weed memberships include: Zostera, seagrass, soft algae, kelp, Cladophora, weed. These data are usually in a SFT theme line. For specific data on seagrasses, refer to the CMP output file.

Hydrographic Bottom Type (HBT) codes can be output. What do they mean?

- Many navies and engineering groups use these codes. The British Admiralty set used here (British Admiralty, 1973) are not very different from U.S. usage within USCG and NOAA. Some users of dbSEABED will find the codes familiar and helpful. Note that the terms at the front of the code are the most significant (abundant). For the clearest plottings, use the no-overlap options in your GIS and plot only the surficial seabed samples of an area.
- The codes output in EXT are either passed direct from naval/engineering HBT codings or re-codings of the lithological types that scientists have described in sediment samples. The codes output in CLC refer only to the gravel, sand, mud, rock and weed components present in a sediment or at a site.

Note: Data for Hydrographic Bottom Type have not been included in this publication.

What does the Roughness code mean?

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

- This is a coded output representing the V:H of the seabed roughness element which is observed with greatest aspect ratio. That feature may be fixed roughness like a cobble, or moveable roughness like ripples. The outputs can only report observed roughness elements, so are influenced by the size scales of samplers and observations.
- The V and H values are the centimeter values of the height and horizontal dimensions written in integer log 2 (base 2). For example "4:6" represents 16 cm height over length scale of 64 cm. Powers <0 are set to zero (i.e., scales <1 cm are not considered). The horizontal length H is the length of expression of a feature, rather than wavelength of repetition. Where a feature is elongated, H is taken normal to elongation (i.e., equals ripple wavelength).

Why do some points plot on the land?

Some of the data collections we used contain data in estuaries, rivers and lakes; others contain data for coastal dunes, beaches, even coal mine pits. We could have culled these data, especially the ones that are obviously not marine. But the accuracy of maps of the national shoreline is not good enough to do a cull based on position of samples alone. And some people will find the dune, river, and lake data useful. Therefore we left all the data in.

Can the same sample or analysis be represented more than once in the data collection?

Yes, when there is a good reason. For instance, the same sample may be described by two different labs or be re-analysed years later. Or an analysis may be expressed in a different way by later work. These duplicate results are all valid, and should be included in the database for mapping. (Note that the duplicate analyses may not carry exactly the same sample names.)

Why are the Kurtosis and Skewness not reported from the database?

These higher-order moments are not only rarer in data sets, but to be accurately reported require a higher standard of data and calculation than lower moments. Statistical moment and graphical measures are also very difficult to reconcile, exacerbating the shortage of available data on which to base outputs.

How does dbSEABED make word data conformable with numerical data?

- dbSEABED software attaches Fuzzy Set Theory memberships to word terms through a look-up dictionary table. The memberships are summed across a description and the totals are output. Some words like "slightly" only adjust the memberships. Other words adjust the character of another, for instance "fine" in "fine sand". For more detail, see Jenkins, 1997.

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

- dbSEABED also recognizes if a word is not in its dictionary, if the meaning is unknown. For example: “zyzgy” is not in the dictionary and will be labeled an ‘unknown’ It also recognizes if a word is neutral for a question. For example: “sediment” is neutral on the question of color.

What if a user doesn't want to use the word-based data for mappings?

Users have a choice: if only numerical analytical data are required, then the ‘extracted’ (*EXT*) form of output should be used; if word data only, then the ‘parsed’ (*PRS*) form. Some users will not want to use the ‘calculated’ (*CLC*) outputs which have a higher level of uncertainty than these, and that is possible too. If a user wants to integrate these to achieve the best possible coverage in a region, then they can be added or telescoped together using two other formats (*ALL* and *ONE* formats respectively; not included in this publication).

Which form of data should I trust more? Numeric-value analyses or word-based descriptions?

There are several issues that users must consider in deciding which form of data to use:

- Issue 1. Descriptive data are more representative of the seabed, the shells, coral, rocks and other items that will not fit in laboratory bottles or analytical instruments but which have a power effect on erodibility, acoustics and habitat. It is also essential for dealing with the creatures of the bottom, burrows and other structures, sediment strength, odors, color, etc.
- Issue 2. To get enough data to make a map with a useful number of points, you will have to use word-based data, which account for 85% of all available data in usSEABED.
- Issue 3. Analytical data are more precise, often precise as to within 2% of the full parameter range. And it is repeatable, given the same instrument unlike descriptions which may vary between observers, even trained ones. However, it may not be accurate for the seafloor, as it usually only deals with the seabed matrix – those materials fine enough to fit inside the instrument tubes / apertures.

How do word-based descriptive data relate to numeric-value analytical data?

- The values of grain size, percent gravel, carbonate, etc., that are calculated from word-based data are based on Fuzzy Set Theory calculations. They are memberships, reduced to a single value. An analysis is also a single realized value of the Fuzzy Set.
- Sometimes word-based and numeric-value results are available for the same sample, and then an inter-calibration of the two forms of output can be made.

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

Why are values of grain size only given to 1 decimal place, and percent sand, etc., given as integers?

Work in the dbSEABED project has established the typical accuracies on measurements of these parameters. Those accuracies, even for careful work are of the order of 2-5% of half the total range of the parameter. The table outputs are designed to be brief to restrict data volumes and make mapping faster and easier and to carry the data at a precision that is appropriate to measurement accuracies.

How are the grain sizes extracted from data?

By a simple reporting of the average grain size, median grain size, Inman Mean, and Folk Graphic Mean grain sizes. A comparison of data sets shows that these reflections of 'central' grain size are not significantly different within usual error bounds on sampling then analysis. (On the other hand, sorting values are significantly different and cannot be combined in the same way.)

Are detailed grain size analyses used to output textural statistics like mean and sorting?

- Yes. The percentages in each fraction are summed to output the percent gravel, sand, mud and clay.
- A weighted average and standard deviation is also calculated, leading to output of central and sorting grain sizes. Standard moment statistics (see Blatt and others, 1980) are used.

Why would a sample that is known to have had a detailed grain size analysis not appear in the output data?

grain size analyses are scanned to check that they are in good order, and some are rejected. An analysis with any phi interval is acceptable, provided it is fine enough to resolve those fractions. Analyses that have a significant weight percent in the finest and coarsest classes are treated as suspect because this implies that the part of the sediment that was analysed probably does not represent the whole sediment.

How are the outputs for the grain size fractions gravel, sand, mud extracted?

Many data sets contain these values (based on Wentworth scale), and in those cases the values are passed through to outputs (EXT). In some cases the data are presented in the form of a detailed grain size analysis – such as at ½ phi intervals. dbSEABED assembles grain size analysis streams into G:S:M fractions by assigning each analysis class to its fraction (or proportioning if the class straddles a fraction boundary).

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

Note: Many grain size analyses techniques range only through sand and mud, and gravel is not analysed. In these cases the database reports the gravel percent as null in outputs.

Are grain size scales other than Wentworth used?

Generally no. However, an option does exist, to report percent mud from values of 'engineering' grain size mud (i.e., finer than #200 Sieve, 75 μ m).

How is one central grain size recovered from diverse data that may present mean, median and other graphical grain sizes?

The database adopts mean moment grain size as the standard for its measure of a sediment's central grain size. Studies show that the median, Inman 'mean', and Folk 'mean' grain sizes fit this quite well for a wide variety of sediments. However mode grain sizes do not, and are not included in 'central grain size'.

How are the central grain size and sorting estimated in the CLC process of the database?

- Many sediments have gravel:sand:silt:clay (g:s:s:c) or gravel:sand:mud (g:s:m) ratios reported, but are without basic grain size statistics. Some users of dbSEABED requested that 'best estimates' of the statistics be made by modeling grain size distributions. This is done by creating a weighted mean and standard deviation from the g:s:s:c and g:s:m ratios of the samples.
- The class statistics underpinning the modeling were arrived at by examining the most common values for single phase sediments in USGS data sets (for example, just silt). The values were for central grain sizes g:s:(m):s:c -3.0: 2.0: (7.0): 5.0: 8.0. Typical values of sorting were also applied: for g:s:(m):s:c 1.4: .0.9: (2.1): 3.0: 5.8. These values may be adjusted in the future.

What is Fuzzy Logic and how does it work?

"Fuzzy Logic" (more properly "Fuzzy Set Theory") allows an object to belong partially to a set. In classical 'Crisp' Set Theory objects are either in a set or not. FST suits words because they are often partial carriers of meaning. For example, 'warm' is partially hot and partially cold. A formal arithmetic for Fuzzy Sets was discovered by Zadeh, (1965). A good reference is Mott and others, (1986).

What is the method of attributing subbottom depths to samples from cores?

Ideally a sample from a core will have a subbottom depth assigned in terms of meters below the sediment surface. However this is not always the case. While this information may have to be left unknown if not given by the original researchers, some limits can be placed if the sampler type is known. For example a Shipek Grab usually has only 5 cm of

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

penetration. Then a subbottom range of 0-0.05 m can be assigned. In cores that may be several meters long, therefore the limitation is less strict.

How are the shear strengths derived?

Shear strength values are obtained either from actual measurements held in the database or are assessed from descriptions that convey lithification or consolidation.

- These measures of strength are accepted: hand penetrometer strength, vane shear strength (maximum, unremolded), cohesion values from triaxial and shear box tests, compressive shear strength at low confining pressures, and cone resistances. Obviously, uncertainties arise not only from measurements themselves but by combining the different methodologies.
- Many descriptions refer to a sediment or rock being “soupy”, “loose”, “soft”, “cohesive” (“firm”), “stiff”, “friable”, “cemented”, “lithified” or “hard”. These terms are indications of the mechanical state of the materials and are transferred to a Shear Strength value in PRS outputs. They are (respectively): 2, 25, 50, 100, 500, 1000, 5000, 8000, 10000.

What is the method for arriving at the Shepard and Folk classes?

The original Shepard (1954) and Folk (1954) ternary classifications have had to be modified in a couple of ways: (a) where the silt and clay breakdown of mud is not available, the silt-clay domains in the classifications are merged under one name; and (b) because these schemes refer only to sediments, an extra class “SOLID” is added for dbSEABED to classify all lithified materials (rock). The aim of (a) was to have a class attached to a greater number of samples. Very few samples in the database have silt:clay specified.

What relationships are used to estimate the Critical Shear Stress values?

Depending on the quality of inputs and nature of the sediment, several relationships are used to predict Critical Shear Stresses. The details and supporting references are given in the onCALCULATION document. The relationships include:

- Where the consolidation is known and substantial, the CSS is set equal to the value of the Shear Strength.
- For unconsolidated granular sediments a compilation using hundreds of experimental results from the field and laboratories suggests a linear relationship with phi grain size.
- For unconsolidated fine-grained sediments (>5phi grain size) where there is information on density or porosity, the CSS is based on the density using published relationships.
- For unconsolidated fine grained sediments without information on density or porosity, a general value is used.

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

<>No regard to the effects of bio-consolidation or bioturbation is made.

How are the porosity values obtained?

- First, values that are in the database are passed to the EXT outputs. In addition, where bulk or dry densities, void ratios, or moisture contents are present, they are transferred to a porosity value. Where this requires a grain density, this may be available, or is assumed to be 2.5 g/cm³ which is suitable for a wide range of materials.
- To predict porosities for unconsolidated sediments we use an empirical relationship between porosity and grain sizes described by Richardson and Briggs' (1993): $AvGrsz = -4.55 + 0.169 * Por$, with R² of 0.81 in the range 0-12 phi. We use the inverted form: $Por = 26.92 + 5.92 * AvGrsz$. These results appear in the CLC outputs. Other unpublished relationships of higher accuracy are available.

What is the basis for the data on compressional (p-wave) sound speeds?

Although there are some direct measurements of compressional sound speeds for areas of the seabed, they are few and far between. dbSEABED reports any direct measurements in the EXT outputs, including the average of those data which recognize velocity anisotropy. One difficulty with sound speeds is that they do depend on the frequency, pressure and temperature of the measurement. The outputs make no allowance for these factors, though conditions of the original measurements are recorded in the underlying raw data of the database.

Several users of dbSEABED requested that estimated sound speeds be provided wherever possible. This is done as follows.

- Where the material is consolidated and a measured porosity is available, then the time average equation is employed (Nafe and Drake (1960)). Consolidation is judged to be when the ShrStrs based on measurement or a parsing of compaction/cementation is > 50kPa, or the porosity is < 33%. These values for constants were adopted: solids and fluid densities 2500 and 1025 kg/m³, respectively; sound speed of solids and fluids 5000 and 1520 m/s, respectively.
- Alternatively, where the material is a loose sediment the velocity is calculated based on the relationships of Richardson and Briggs (1993): $Vb/Vw = 1.18 - 0.034 * Mz + 0.0013 * Mz^2$ with an R² of 0.82, where Vb/Vw is the ratio of seafloor sediment and bottom water p-wave velocities. For an absolute output velocity, this result is combined with the bottom water velocity (assumed 1520 m/s). Other unpublished relationships of higher accuracy are available. The method makes no allowance for the consolidation of a sediment below the seafloor, for example in a core.

How are the sorting values arrived at?

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

- Values of moment sorting are simply passed through to output. Graphical measures of sorting such as those of Inman and Folk are not generally compatible with moment measures and at present do not contribute to outputs.

Carbonate values: how are they arrived at? And Organic Carbon?

- Carbonate measurements that have been made on the complete sediment are simply reported in the EXT outputs. Many sediments have had carbonate determined on just the sand or mud fractions: those results are not output because they do not represent the whole sediment.
- Carbonate values are also output from the parsing process. The carbonate parts of sediment are summed in quantity, as are the non-carbonate and unknown carbonate parts (like 'mud'). If the sum of unknowns is less than 5% then a carbonate membership is reported in PRS outputs. The accuracy of this result is not as good as for analysis data.
- Organic carbon data are treated the same way. However, because OC values are typically small in sediments (~1%), the accuracies in the PRS results are not good and should be treated with caution at this stage.

How does dbSEABED hold the basic data?

- Datasets for processing are held in structured documents that are called "Data Resource Files" (DRF). They set out data in the form of a written geological core-log, in a tree structure that nests the data according to expedition, the sample site, the sample, and finally phase inside the material. This format is processed by programs to produce other formats that are compatible with GIS and RDB.

Why does dbSEABED hold its underlying data in documents rather than a relational database like Oracle?

This is one of the most successful features of dbSEABED.

- First, it is more efficient to bring in data sets from geologists in the form of written core-logs rather than tables, especially as tables and sub-tables. This is how dbSEABED has been able to quickly absorb a million data sites.
- Second, it is not possible (easily, inexpensively) to do some sedimentological operations on the data in software like Oracle. For example, parsing the word-based descriptions.
- Third, in the documents, it is possible to have data which was obtained at the same time (such as replicates, suites of analyses) to be kept right next to each other. The same is true for the metadata; a person can read the data and the metadata about it easily and at the same time.

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

- Fourth, it keeps the data very faithfully in its original form. Relational databases require that data be shuffled and diced, and that sets of rules are obeyed before the data can be entered. Data are kept faithful to that of the original observers.
- Relational Databases are just one type of data structure. There is no reason why dbSEABED shouldn't be able to generate others too, like XML, and Spreadsheets and GIS tables.

What is involved in importing data sets into dbSEABED?

- This depends on the structure and idiosyncrasies of the original data. Many data sets are simple to import into the Data Resource File (DRF) format. Others have proven to be most difficult. It is usually done by arranging data items by column in EXCEL according to a template that sets out the field locations for each data item (such as latitude, sampler type, phi grain size, and Munsell color code). Different types of data are divided between themes such as geotechnical, texture, or color. Some themes allow for a sample to be from the subbottom (as in a core). Others such as seafloor type which is dedicated to surficial sediment surveys such as those by divers and remote vehicles, treat only the seafloor surface.
- Although dbSEABED tries to hold data as close as possible in its original terms, this is not always possible. In the case of numeric data items, if a simple conversion of units is desirable – for instance from cm into m – then that can be done at data import stages. In the case of word descriptions, it would be unwieldy to hold or process the descriptions in original long words and words are therefore converted to abbreviations. Note that this does not change the terms, it is NOT a re-classification. Terms such as “lithothamnion”, “scattered”, “low water content” are abbreviated to “lthmn”, “scatrd” and “lo_watr_cntnt”. This speeds processing, allows importing staff to check the data for spelling errors, allows us to use foreign languages (for example, “sabb(IT) for “sabbiosi”), and permits homonyms to be distinguished (eg: dens_ab” and “dens” for abundance and geotechnical densities). It also allows the data to be read more easily on computer screens.

How is metadata treated?

Metadata are treated at several levels. First, over-arching metadata have been compiled for the processed outputs of dbSEABED (see “http://instaar.colorado.edu/~jenkinsc/dbseabed/resources/db9_MetadataFGDC.txt”). Second, the USGS and others have compiled metadata for individual data sets as they are entered into dbSEABED. Third, metadata describing measurement techniques, features and problems in the data, edits that we have made in dbSEABED, who collected and analysed the data, etc is held in direct association with the data in the DRF files. This most detailed metadata are best viewed in the RDB structure, which also puts out reports when data values are beyond plausible limits, for example a grain size of 14 phi.

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

Why are sediment descriptions and biological names held as abbreviations?

To make the parsing (and dictionary lookup) computationally faster, to make long descriptions shorter and more readable, to give flexibility in handling homonyms, and to better distinguish active data from metadata in the structured documents.

Where can I find more information on the Munsell color code?

- Refer to the Geological Society of America “Color Rock Chart,” or to the company GretagMacbeth which is the modern-day custodian of Munsell's color technologies.

How are the Munsell color codes derived?

This process is fully described in Jenkins (2002). Munsell codes are explained in a publication of the Geological Society of America (Goddard and others, 1951). The essential steps in the treatment of color data are as follows:

- If data are available in Munsell Code form then that is transferred to outputs.
- Where a color description like “light greenish gray” is available, then that is parsed by a weighted summing of the hue (color), chroma (color intensity) and value (greyness) of the terms. From these components a Munsell Code is formed. The process is calibrated using the GSA scales and others.
- Outputs are stepped at intervals of 5 in hue, and 3 in chroma and value. Otherwise the full range of >3240 possible natural codes would not be mappable. This stepping arrangement reduces the number to about 40, as seen in the ready-to-use legend that is available with this publication. That legend was constructed by adjusting the RGB values of the symbols to match their Munsell colors as seen on a computer screen.

How can I map the coded information on Color and Roughness in a GIS?

- Load the ArcView legends “munsell3.avl” or “rgh_pt.avl” which are available with the database. ArcView legends may be imported into ArcGIS.
- To make your own legends for other applications, employ a classification that uses a “unique value” process.

How can I use the Critical Shear Stress in a practical application?

- The CSS and grain size are essential components for calculating the Shields criterion of sediment erosion, between the Shields Parameter and the Grain Reynolds Number. Those parameters can be calculated once the flow characteristics, fluid velocities and densities, etc are known. See standard textbooks on sediment erosion for more information (for example, Soulsby (1997)).

Frequently Asked Questions (FAQs) about dbSEABED *in*: Buczkowski, B.J., Reid, J.A., Jenkins, C.J., Reid, J.M., Williams, S.J., and Flocks, J.G., 2006, usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and U.S. Virgin Islands) Offshore Surficial Sediment Data Release: U.S. Geological Survey Data Series 146, version 1.0. Online at <http://pubs.usgs.gov/ds/2006/146/>

FAQ References:

Blatt, H., Middleton, G.V. and Murray, R.C., 1980, Origin of sedimentary rocks (2d ed): Prentice-Hall, Inc., 782 p.

British Admiralty, 1973, Chart 5011, Symbols and abbreviations used on Admiralty Charts (new series): Taunton, UK, The British Hydrographic Office.

Folk, R.L., 1954, The distinction between grain size and mineral composition in sedimentary rock nomenclature: *Journal of Geology*, v. 62 no. 4, p. 344-359.

Folk, R.L., 1974, The petrology of sedimentary rocks: Austin, Tex., Hemphill Publishing Co., 182 p.

Goddard E.N., Trask, P.D., de Ford, R.K., Rove, O.N., Singewald, J.T. and Overbeck, R.M., 1951, Rock color chart: Geological Society of American.

Mott, J.L., Kandel, A., and Baker, T.P., 1986, Discrete mathematics for computer scientists and mathematicians (2d ed.): Reston Publishing Company, 751 p.

Nafe, J.E. & Drake, C.L., 1960, Physical properties of marine sediments, *in* Hill, M.N., ed., *The Sea*, v. 3: New York, Wiley, p. 794-815.

Poppe, L.J., Eliason, A.H., Fredericks, J.J., Rendigs, R.R., Blackwood, D., and Polloni, C.F., 2000, Grain size analysis of marine sediments: methodology and data processing, *in* Poppe, L.J., and Polloni, C.F., eds., USGS East-Coast sediment analysis; Procedures, database, and georeferenced displays: U.S. Geological Survey Open-File Report 00-358, CD-ROM.

Richardson, M.D., and Briggs, K.B., 1993, On the use of acoustic impedance values to determine sediment properties: *Proceedings of the Institute of Acoustics*, v. 15, p. 15-24.

Shepard, F.P., 1954, Nomenclature based on sand-silt-clay ratios: *Journal of Sedimentary Petrology*, v. 24, p. 151-158.

Soulsby, R., 1997, Dynamics of marine sands: London, UK, Thomas Telford, 249 p.

Zadeh, L.A., 1965, Fuzzy sets: *Information and Control*, v. 8, p. 338-353.