

# Classification of geological data as a key to optimizing the preparation of CDI and ODV in the Geo-Seas project

Georgij Konshin, University of Latvia (Latvia), [georgij@lu.lv](mailto:georgij@lu.lv)  
Valdis Seglinsh, University of Latvia (Latvia), [valdis.seglins@lu.lv](mailto:valdis.seglins@lu.lv)

In this report consider some topical issues of implementation the project GeoSeas in order to optimize the preparation of records CDI and ODV. The main reason that prompted us to prepare this report is that the existing guidelines and instructions for all their quality have one drawback: the examples tend to relate to the field of oceanography and is not always understood by geologists. In other words, over the Geo Seas dominates the experience of its predecessor SeeDataNet. Therefore it is reasonable to explain the problems we encountered in simple examples and we hope that it will be useful for geological data description and probably is necessary to determine main questions solved during the project progress.

**1. Bottom sediments - main pilot object.** The bottom sediments (BS) are the substantial object in the Latvian marine geology archives. As a result of geological survey of the Baltic Sea a significant number of geological samples or cores were collected. Their capacity usually ranges from 0.0 to 20-30 cm, rarely reaching several meters. Most of bottom sediment samples were selected through sediment grabs, rarely with a vibrocorer or with drilling. In the first case, the samples do not have the form of core and layered descriptions. However, in other cases, samples were collected as cores and have a simplify sediment (lithologic) description. Samples of bottom sediment were taken at 3,459 stations. The acquired samples were further analyzed in the laboratory using different analyses such as grain size analysis, chemical analysis, spectral analysis, etc. The results of each of these analyses have stored in a separate table of local database Geobank.mdb referring to the same metadata of the field sampling.

**2. Data set definition.** We considered that the separate data set should relate to one station and should coincide with same data set in a local centre where we can download them. The separate dataset must include three groups of metadata. The first group includes the metadata that relate to the CDI XML format. This metadata define data set identification and give answers to the following basic questions e.g. Where? When? What? How? Who? Where to find data? Station / Cruise data? Other relevant information? Other relevant services? The second group includes the metadata that relate to the conditions of sampling. The third group includes the metadata that relate to the results of analysis of samples.

**3. The data set name** consists of a combination of three elements: abbreviation of the geological object type (e.g., BS - Bottom sediments); abbreviations of analyze: grn (granulometric), ch (chemical), sp (spectral), min (mineralogical) ptr (petrographical) or it combinations (e.g., grnsp); station's number throughout local database (e.g., 1003). So get like BSgrn1003. The names of Dataset\_ID for the identification data set consist of the last two elements, for example, grn1003.

**4. Creating the Test database for approval the methods of CDI records and ODV files preparation** To create a pilot database must be clear about the basic principles and mechanism of action of Geo-Seas as an information retrieval system. As can be seen, an important feature of GeoSeas is the use of such specific means as CDI and ODV. It should be emphasized that the use of ODV as a data transport format is one of the characteristics of GeoSeas, what distinguishes this system from databases EU-Seased (Eurocore, Euromarsin. Euroseismic), where access to samples and other data sets is determined by negotiation between the user and the repository where the sample is stored..

**5. Data set classification.** It is clear that in determining the composition of test database is necessary to consider how it presents geological data A. Sampling factors/conditions: A1. At the station was taken only one sample usually by sediment grab. It not looks like a core. A2. At the station were taken a few samples usually by vibrocorer or drilling. Samples were taken at different depths below the bottom of the sea. It looks like a core, for which a core logs / profile was made (lithology). B. Laboratory studies. B1. The sample was studied for one kind of analyze (e.g. spectral analyze). That is **Simple data set**. B2. The Sample was studied for several kind of analyze (e.g. spectral and granulometric). That is **Complex data set**. Simple data set are divided into 6 groups: BSgrn, BSch, BSmin, BSpt, BSsp with the data of granulometric, chemical, mineralogical, petrographical, spectral analysis, respectively. Complex data set are divided into 9 groups: BSgrnsp, BSgrnch, BSgrnmin, BSsptr, BSgrnchsp, BSgrnptrsp, BSgrnminch, BSgrnminsp, BSgrnspchmin with the data of two or three or even four kinds of analysis of the six possible. Based on the proposed classification, we created a test database consisting of 4 data sets which cover geological data sets different in content and structure. 5.1. A1&B1 (e.g., BSsp). 5.2. A1&B2(e.g., BSgrnsp). 5.3. A2&B1(e.g., BSch). 5.4. A2&B2(e.g., BSgrnsp).

Data sets	Sampling factors	Laboratory studies	Data name
1	A1 one sample usually by sediment grab	B1 The sample was studied for one kind of analyze	BSsp
2	A1 one sample usually by sediment grab	B2 The Sample was studied for several kind of analyze	BSgrnsp
3	A2 few samples usually by vibrocorer	B1 The sample was studied for one kind of analyze	BSch
4	A2 few samples usually by vibrocorer	B2 The Sample was studied for several kind of analyze	BSgrnsp

**6. ODV main principles.** One of the principles of putting geological data in ODV format is, it is more convenient to separate in-situ measurements and ex-situ measurements (such as laboratory analyses) into separate ODV files. So the solution would be to have only one CDI file, describing the \*field\* measurement, and two ODV files: one describing the lithology in the field and the other one containing the parameters measured in the lab. These two files have the same structure but different content. These files will be zipped and this will be the data file that is listed in coupling.txt.

### 7. ODV classification

Data sets	Sampling factors	Laboratory studies	ODV type
BSgrn12	A1	B1	<b>ODV1</b> on the station was selected one sample which studied using a single analysis
BSgrnsp506	A1	B2	<b>ODV2</b> on the station was selected one sample which studied using two analysis
BSch993, BSsp3 BSmin25, ptr1341,	A2	B3	<b>ODV3</b> on the station was selected several samples which studied using one analysis

In conclusion it should be noted that the proposed classification is deliberately allowed to approach the method of preparation CDI and ODV during the Geo-Seas project.