Technical Report

Interoperability between MARLIN and CRITTERBASE

Authors: Paul Kloss, Dr. Jennifer Dannheim

(Technical report of the project „ANsätze zur Kostenreduzierung bei der ERhebung von Monitoringdaten für Offshore Vorhaben“ (ANKER), Arbeitspaket BENTHOS; FKZ 0325921)

Last update: 03. June 2020

Citation: Kloss P, Dannheim J (2020). Database interoperability between MARLIN and CRITTERBASE. Technical Report of the project ANKER FKZ 0325921, pp. 9.

Alfred-Wegener-Institut, Helmholtz-Zentrum
für Polar- und Meeresforschung (AWI)
Am Handelshafen 12
27570 Bremerhaven

1. Aim

The aim is to guarantee data exchange between different information systems in order to use data of different systems for large-scale and long-term biological analysis for a variety of purposes and information. Thus, the goal is to create a data transfer between the information systems MARLIN and CRITTERBASE that should be kept as simple and uncomplicated as possible. For this purpose, an import/export module based on Excel is under development.

2. Introduction

Sustainable management of marine resources and conservation of species and habitats requires sound scientific knowledge on species distribution in space and time, as well as natural and human-induced changes of populations. This understanding of the systems properties can only be achieved by having a broad view on the benthic ecosystem, i.e. looking at large-scale and long-term trends. Broadening our view increase our understanding on how the benthos responds to pressures, how we determine the level of change and ultimately, whether there are any effects (Halpern et al. 2008, Ban et al. 2010). Thus, the tasks for applied science have changed fundamentally and call for a rethinking in the scientific application of large-scale and long-term data sets by multidisciplinary research in order to give scientific advice. In science, geo-statistical analysis and prognostic models experience progressive attention and implementation to explain natural benthic variability or to what extend planned activities might affect the sustainability of the benthos. However, ecosystem management is often confronted with fragmented information on the distribution of marine species and habitats, because marine environments are more difficult to access and to monitor compared with terrestrial ecosystems (Robinson et al. 2011). While decision processes and regulations are rather implemented at large and long-time scales (e.g. Marine strategy framework directive, Habitat directive, Marine Spatial Planning), research studies and effect assessments are linked to specific projects or impact assessments with a limited temporal and spatial coverage (Floen et al. 1993, Vanden Berghe et al. 2009). Isolated datasets from projects are thus of limited use for the interpretation of large-scale and long-term benthic effects in the framework such as climate change or directives at a broad national or European level (Costello & Vanden Berghe 2006, Vanden Berghe et al. 2009). Individual studies are restricted in the amount of data they can generate, but by combining the results from many studies, massive databases can be crated that make analysis on a much-enhanced scale possible (Grassle 2000).

Over the past years, several information systems have been developed due to the increasing availability of environmental data (Brown et al. 2011) and the advances in computer

technologies. Examples are the Ocean Biogeographic Information System (OBIS), the data integration component of the Census of Marine Life (CoML), the integrated database on soft-bottom macrobenthos MacroBen within the MarBEF (Marine biodiversity and Ecosystem Functioning) initiative, INSPIRE as an infrastructure for spatial information in the European Community and EMODnet as the European Marine Observation and data network. On a national level, the MDI-DE (Marine Daten Infrastruktur, Deutschland) and GeoSeaPortal are portals that joins all German coastal data and information in order to fulfill the scientific needs of improving EU- guidelines such as MSFD and INSPIRE.

However, these information systems are often based on meta-data, provide information on benthos as pre-fabricated products (in forms of table, graphs or maps), or do not provide tools to analyse raw data (i.e. raw data provision only). Further, data information systems are, more often than not, either initiated by regulatory authorities of marine systems, i.e. European or national authorities, or by international scientific initiatives. Either way, this leads to a lack of data from different authorities.

Here, we have developed data information systems, MARLIN of BSH and CRITTERBASE of AWI, which are able to exchange data in a harmonised, plausibility-checked and high-quality assured way. Thus, information on the benthic system can be provided on the fly based on these combined high-quality standardised data by tools of the information system. For the first time, this offers the opportunity for the German Bight to provide information at a high resolution, larger-areas and longer time scales by the interoperability of systems. Further, this allows for cumulative impact assessments of e.g. offshore wind farms and for disentangling natural variability from anthropogenic changes in the benthos based on a sound and detailed scientific database.

3. Implementation in MARLIN

Within MARLIN there are multiple import options using EXCEL files. A separate import and export path is being established so that MARLIN can handle CRITTERBASE data.

3.1. Creating an Excel template

In MARLIN, an Excel import template via graphical user interface was created. Each entry is described in detail. Furthermore, a plausibility check entry is created automatically for each

entry, which can be configured later on. Figure 1 shows the import configuration dialog; the column „PROJECT“ in worksheet „Hol-Beschreibung“ is selected.

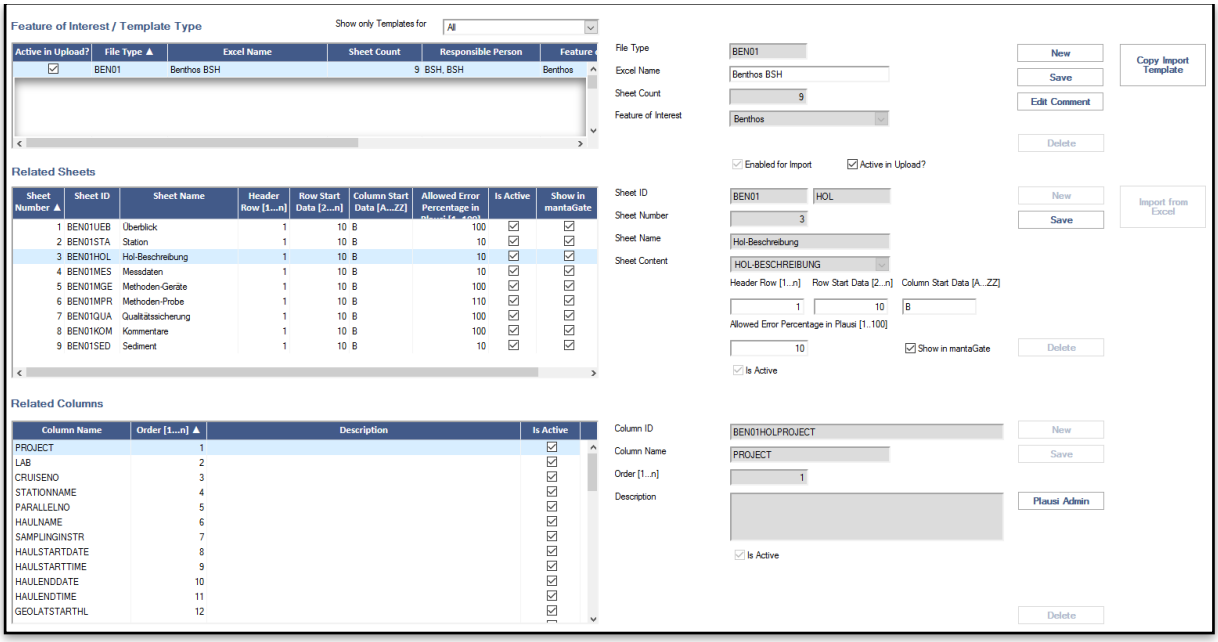


Figure 1: Screenshot of import configuration dialog in MARLIN

Figure 2 shows a configuration dialog to edit a plausibility check. The plausibility test for column „SAMPLINGINSTR“ in worksheet „Hol-Beschreibung“ was selected here. The following options can be selected here very easily.

- Is the specification of the attribute obligatory?
- The data type of the attribute
- Various size specifications for different formats (e.g. string length, date range, number interval)
- Must the recordable value of this attribute be findable in a catalog?

Furthermore, the answers for violations of each of these rules can be defined.

Predefined SQL test queries can also be selected and activated from a drop-down menu. SQL test queries can also be stored in the system. However, this does no longer work simply via the user interface but requires a direct adaptation of SQL scripts to the corresponding storage locations on the storage device. In other words: this is hard coding work.

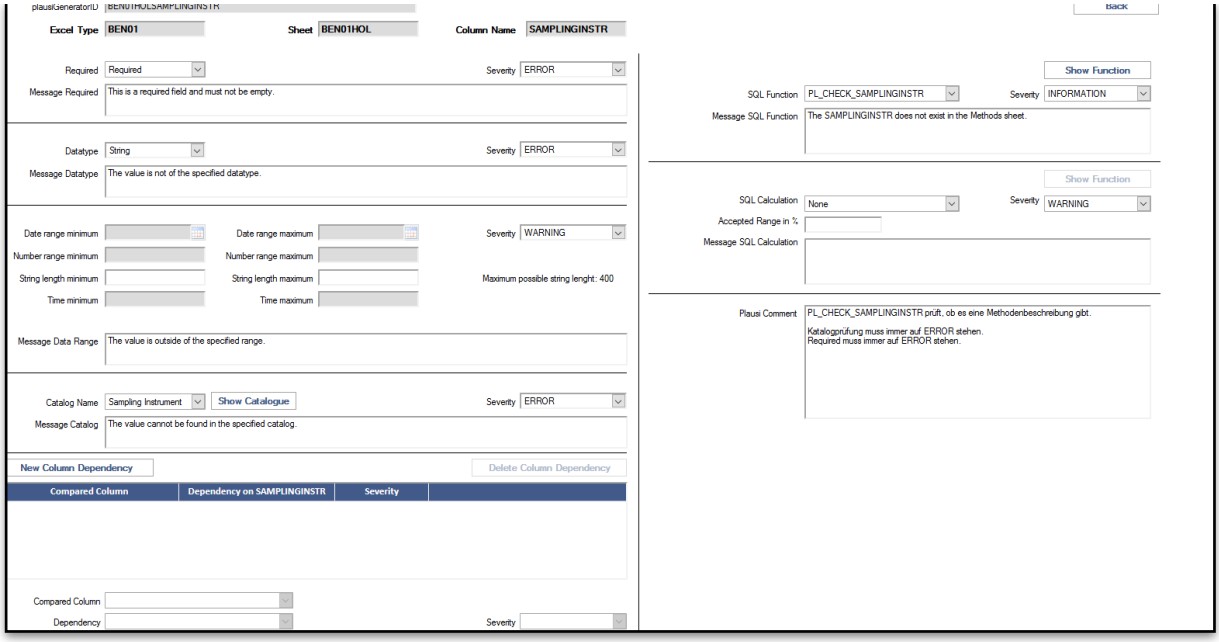


Figure 2: Configuration options for a plausibility test.

3.2. Import from CRITTERBASE

Data delivery

The actual data delivery takes place in the DMZ (demilitarised zone) in the internet. Figure 3 shows the interface where the import will take place in MARLIN.

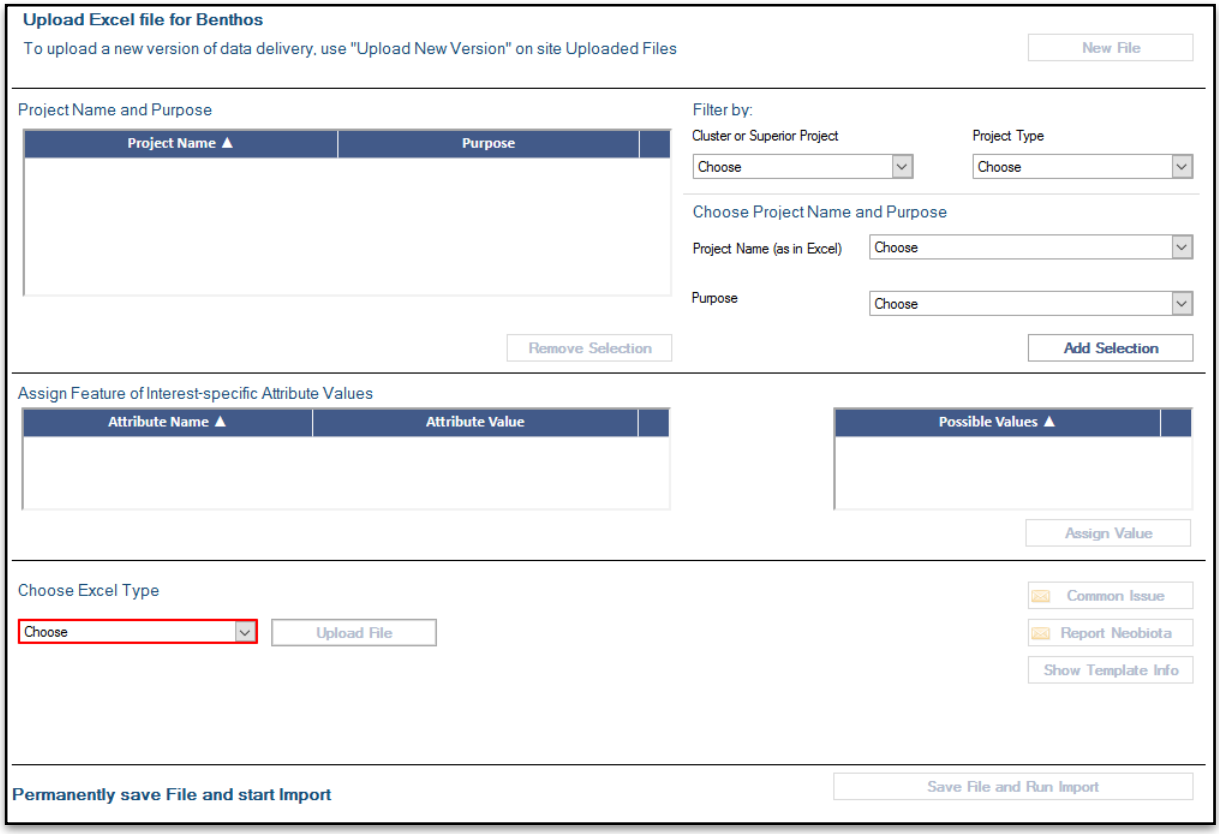


Figure 3: Dialog in which the CRITTERBASE import will be located in future.

Import Magic

The import logic cannot simply be generated and was programmed in a dedicated way. The area in which this is implemented is called MARLIN Import Magic. The specific conversions and adjustments of the incoming data must then be carried out in this area. This also represents more difficult development work.

3.3. Export to CRITTERBASE

The export into an Excel table is conceivable as a separate use case and represents a separate product in MARLIN. Figure 4 shows the place in the systems interface structure where this export will be integrated.

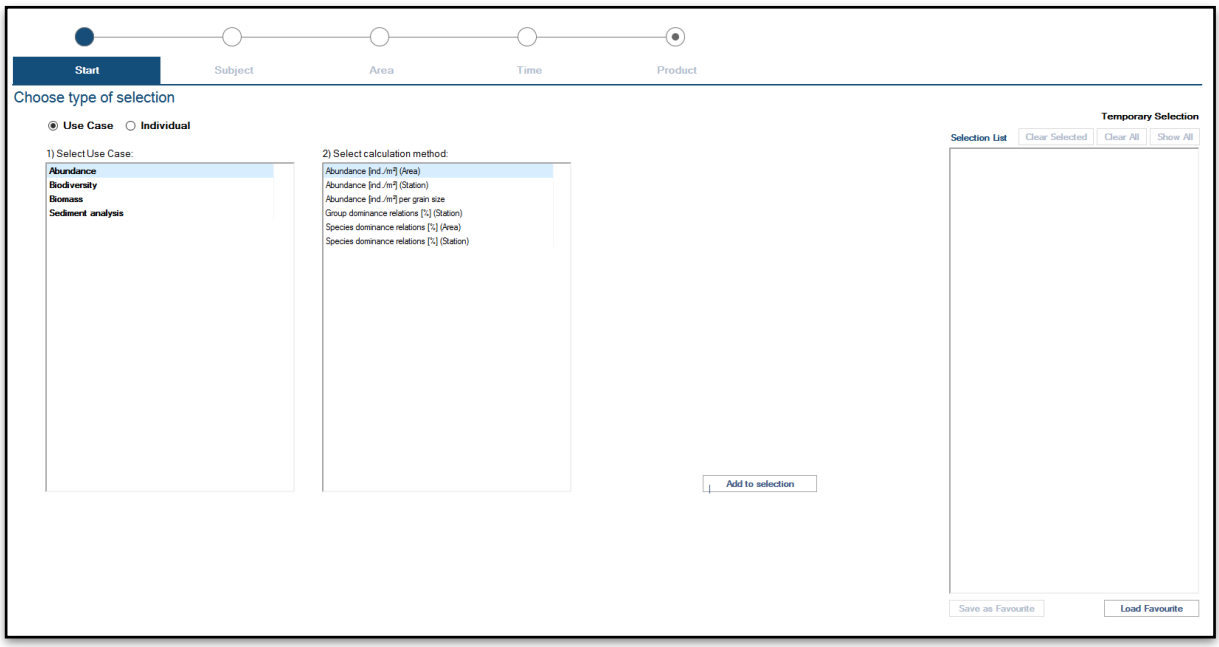


Figure 4: Dialog in which the export to CRITTERBASE will be located in future.

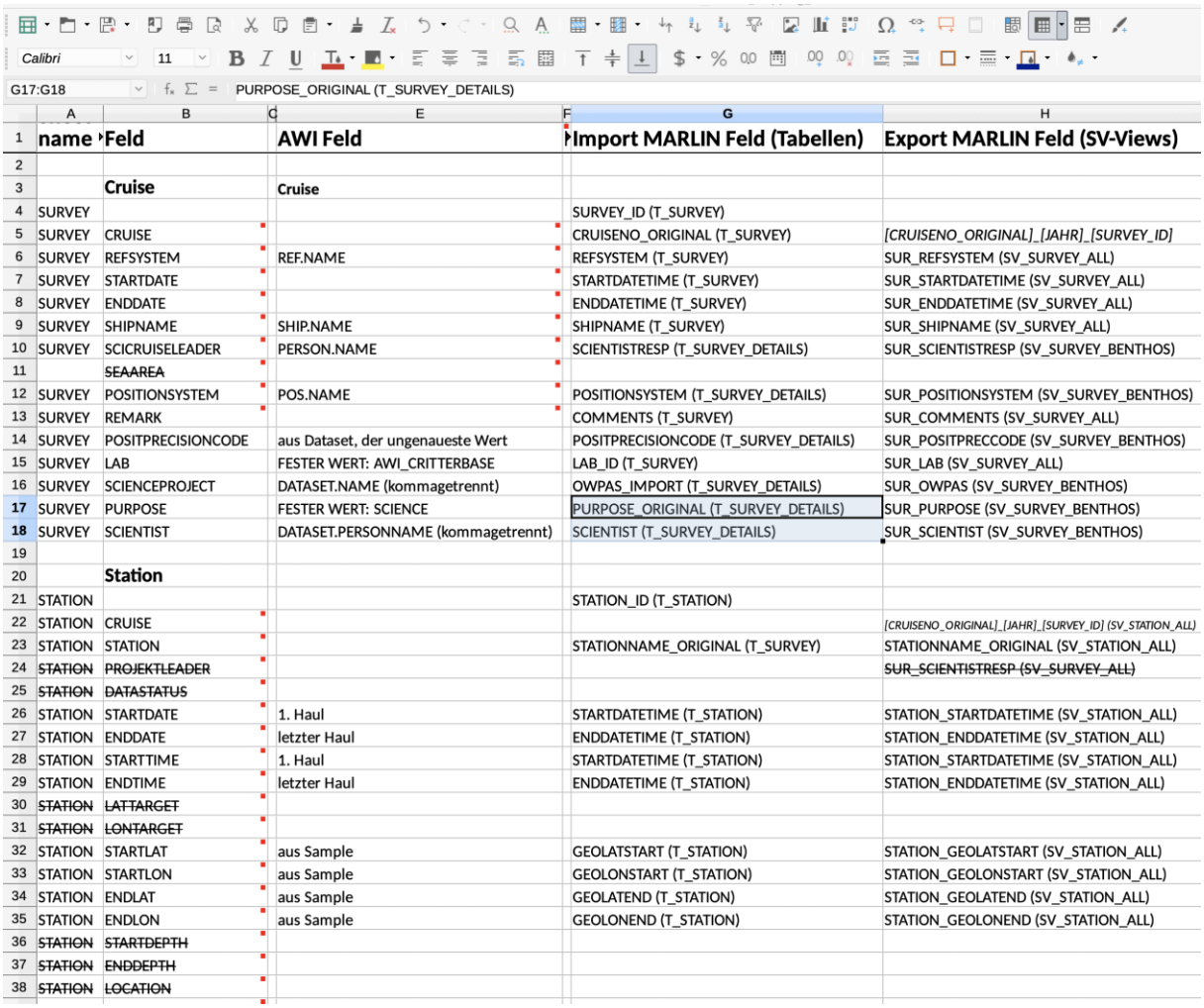
4. Technical descriptions

The databases of the MARLIN and CRITTERBASE have to be linked to guarantee continuous data exchange. For this purpose, the two data models of CRITTERBASE and MARLIN were initially compared and harmonized (figure 5).

A first draft of a transport format has been developed into which each system will import and export data.

Disharmonies are addressed and, if possible, resolved in the import or export routine in the respective target system. It is therefore up to each system to create the export and import modules itself. If differences occurred, they have been resolved with as little effort as possible.

The localisation of the coding work depends on the effort to solve the problem: If, for example, an import problem on CRITTERBASE side is much easier to solve on MARLIN export side, coding work was carried out at the MARLIN export side and vice versa. Thus, the least possible and reasonable effort was always considered and coded.



G17:G18 PURPOSE_ORIGINAL (T_SURVEY_DETAILS)

	A	B	E	G	H
1	**name**	**Feld**	**AWI Feld**	**Import MARLIN Feld (Tabellen)**	**Export MARLIN Feld (SV-Views)**
2					
3		**Cruise**	**Cruise**		
4	SURVEY			SURVEY_ID (T_SURVEY)	
5	SURVEY	CRUISE		CRUISENO_ORIGINAL (T_SURVEY)	*[CRUISENO_ORIGINAL]_[JAHR]_[SURVEY_ID]*
6	SURVEY	REFSYSTEM	REF.NAME	REFSYSTEM (T_SURVEY)	SUR_REFSYSTEM (SV_SURVEY_ALL)
7	SURVEY	STARTDATE		STARTDATETIME (T_SURVEY)	SUR_STARTDATETIME (SV_SURVEY_ALL)
8	SURVEY	ENDDATE		ENDDATETIME (T_SURVEY)	SUR_ENDDATETIME (SV_SURVEY_ALL)
9	SURVEY	SHIPNAME	SHIP.NAME	SHIPNAME (T_SURVEY)	SUR_SHIPNAME (SV_SURVEY_ALL)
10	SURVEY	SCICRUISELEADER	PERSON.NAME	SCIENTISTRESP (T_SURVEY_DETAILS)	SUR_SCIENTISTRESP (SV_SURVEY_BENTHOS)
11		~~SEAAREA~~			
12	SURVEY	POSITIONSYSTEM	POS.NAME	POSITIONSYSTEM (T_SURVEY_DETAILS)	SUR_POSITIONSYSTEM (SV_SURVEY_BENTHOS)
13	SURVEY	REMARK		COMMENTS (T_SURVEY)	SUR_COMMENTS (SV_SURVEY_ALL)
14	SURVEY	POSITPRECISIONCODE	aus Dataset, der ungenaueste Wert	POSITPRECISIONCODE (T_SURVEY_DETAILS)	SUR_POSITPRECCODE (SV_SURVEY_BENTHOS)
15	SURVEY	LAB	FESTER WERT: AWI_CRITTERBASE	LAB_ID (T_SURVEY)	SUR_LAB (SV_SURVEY_ALL)
16	SURVEY	SCIENCEPROJECT	DATASET.NAME (kommagetrennt)	OWPAS_IMPORT (T_SURVEY_DETAILS)	SUR_OWPAS (SV_SURVEY_BENTHOS)
17	SURVEY	PURPOSE	FESTER WERT: SCIENCE	PURPOSE_ORIGINAL (T_SURVEY_DETAILS)	SUR_PURPOSE (SV_SURVEY_BENTHOS)
18	SURVEY	SCIENTIST	DATASET.PERSONNAME (kommagetrennt)	SCIENTIST (T_SURVEY_DETAILS)	SUR_SCIENTIST (SV_SURVEY_BENTHOS)
19					
20		**Station**			
21	STATION			STATION_ID (T_STATION)	
22	STATION	CRUISE			*[CRUISENO_ORIGINAL]_[JAHR]_[SURVEY_ID] (SV_STATION_ALL)*
23	STATION	STATION		STATIONNAME_ORIGINAL (T_SURVEY)	STATIONNAME_ORIGINAL (SV_STATION_ALL)
24	~~STATION~~	~~PROJEKTLEADER~~			~~SUR_SCIENTISTRESP (SV_SURVEY_ALL)~~
25	~~STATION~~	~~DATASTATUS~~			
26	STATION	STARTDATE	1. Haul	STARTDATETIME (T_STATION)	STATION_STARTDATETIME (SV_STATION_ALL)
27	STATION	ENDDATE	letzter Haul	ENDDATETIME (T_STATION)	STATION_ENDDATETIME (SV_STATION_ALL)
28	STATION	STARTTIME	1. Haul	STARTDATETIME (T_STATION)	STATION_STARTDATETIME (SV_STATION_ALL)
29	STATION	ENDTIME	letzter Haul	ENDDATETIME (T_STATION)	STATION_ENDDATETIME (SV_STATION_ALL)
30	~~STATION~~	~~LATTARGET~~			
31	~~STATION~~	~~LONTARGET~~			
32	STATION	STARTLAT	aus Sample	GEOLATSTART (T_STATION)	STATION_GEOLATSTART (SV_STATION_ALL)
33	STATION	STARTLON	aus Sample	GEOLONSTART (T_STATION)	STATION_GEOLONSTART (SV_STATION_ALL)
34	STATION	ENDLAT	aus Sample	GEOLATEND (T_STATION)	STATION_GEOLATEND (SV_STATION_ALL)
35	STATION	ENDLON	aus Sample	GEOLONEND (T_STATION)	STATION_GEOLONEND (SV_STATION_ALL)
36	~~STATION~~	~~STARTDEPTH~~			
37	~~STATION~~	~~ENDDEPTH~~			
38	~~STATION~~	~~LOCATION~~			

Figure 5: Excerpt of mapping between MARLIN and CRITTERBASE

A basic framework for import/export is in development in CRITTERBASE. It will handle different input formats as far as possible generically. A specific MARLIN-CRITTERBASE workflow based on the actual mapping is in development, too.

5. Results and discussion

Many mapping features between the data models of MARLIN and CRITTERBASE have already been addresses, but the following special features still need to be handled.

Project assignment

Projects are assigned to a cruise in MARLIN and not in subsets as in CRITTERBASE. Subsets subdivide the metadata framework in a much finer granularity. This circumstance

can lead to problems, especially when importing data into MARLIN. In general, the definitions in MARLIN and CRITTERBASE need to be further clarified and refined.

No subset level in MARLIN

CRITTERBASE uses subsets to provide different evaluations of the same sample. This concept does not exist in MARLIN as it is not necessary, but has to be addressed for interoperability of the systems.

Device Catalogs

The equipment catalogues of the systems are underway being harmonized. This is important, for example, as corresponding device types have to be cross checked and marked in MARLIN and CRITTERBASE so that data transfer is possible.

Attributes missing on counter side data model

Some attributes cannot be mapped directly to each other and require a clearly defined conversion convention. For example, CRITTERBASE stores the air temperature per station (TEMP_AIR), while MARLIN stores this data in hauls. When importing data from MARLIN into CRITTERBASE, the value in CRITTERBASEs station table has to be determined from the average values of all corresponding values of all connected hauls in MARLIN Excel sheets.

6. Literature

Ban NC, Alidina HM, Ardron J A (2010). Cumulative impact mapping: advances, relevance and limitations to marine management and conservation, using Canada's Pacific waters as a case study. Marine Policy, 34: 876–886.

Brown CJ, Smith SJ, Lawton P, Anderson JT (2011). Benthic habitat mapping: a review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. Estuarine, Coastal and Shelf Science, 92: 502–520.

Costello MJ, Vanden Berghe E (2006). 'Ocean biodiversity informatics': a new era in marine biology research and management. Marine Ecology Progress Series, 316: 203–214.

Floen S, Gjøsæter H, Korneliussen R, Sagen H, Thorvaldsen P, Wennevik V (1993). An integrated database for marine research. ICES CM 1993 (D:23 Sess.T.)

Grassle JF (2000). The Ocean Biogeographic Information System (OBIS): an on-line, worldwide atlas for accessing, modelling and mapping marine biological data in a multidimensional geographic context. Oceanography (Wash DC) 13: 5–9

Halpern BS, Walbridge S, Selkoe KA, Kappel CV, Micheli F, D'Agrosa C, Bruno JF, et al. (2008). A global map of human impact on marine ecosystems. Science, 319: 948–952.

Robinson LM, Elith J, Hobday AJ, Pearson RG, Kendall BE, Possingham HP, Richardson AJ (2011). Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. Global Ecology and Biogeography, 20: 789–802.

Vanden Berghe E, Claus S, Appeltans W, Faulwetter S, Arvanitidis C, Somerfield PJ, Aleffi IF, et al. (2009). MacroBen integrated database on benthic invertebrates of European continental shelves: a tool for large-scale analysis across Europe. Marine Ecology Progress Series, 382: 225-238.