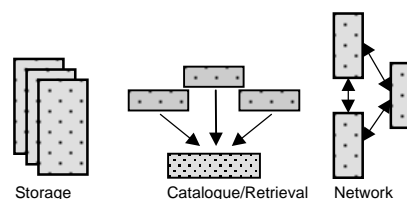


## DATA HANDLING AND THE ANTICIPATED IMPLEMENTATION EFFORT



This annex entails a more detailed overview of the functionality, advantages and implementation effort of the DEMOWAD Data Handling Unit, which is defined by the DEMOWAD data handling project (see "Final Report on the Definition Phase", October 1996).

The prototype test of the Data Handling Unit identified four A-parameter data classes according to its implementation effort (biological, chemical, habitat and general data). The main cost factors of the technical implementation were identified as:

- making data available on the national level and
- making data available on the trilateral level.

With this classification and with a national inventory of TMAP data a rough estimation of implementation effort of A-parameter was assembled.

Derived hereof, an estimation of the requirements and future personnel and technical costs (on the national and trilateral level) of the implementation of the trilateral data exchange system was made.

### Contents:

- I. DESCRIPTION AND MAIN CHARACTERISTICS OF THE DEMOWAD DATA HANDLING UNIT
- II. CLASSIFICATION OF A-PARAMETER DATA
- III. STATUS AND AVAILABILITY OF TMAP DATA ON THE NATIONAL LEVEL
- IV. MAKING TMAP DATA AVAILABLE ON THE TRILATERAL LEVEL
- V. FURTHER IMPLEMENTATION OF THE TRILATERAL DATA HANDLING
- VI. REQUIREMENTS AND EFFORTS OF A TRILATERAL DATA EXCHANGE

**I. DESCRIPTION AND MAIN CHARACTERISTICS OF THE DEMOWAD DATA HANDLING UNIT**

At the present status, the following technical problems have to be solved first when using the monitoring data for the TMAP:

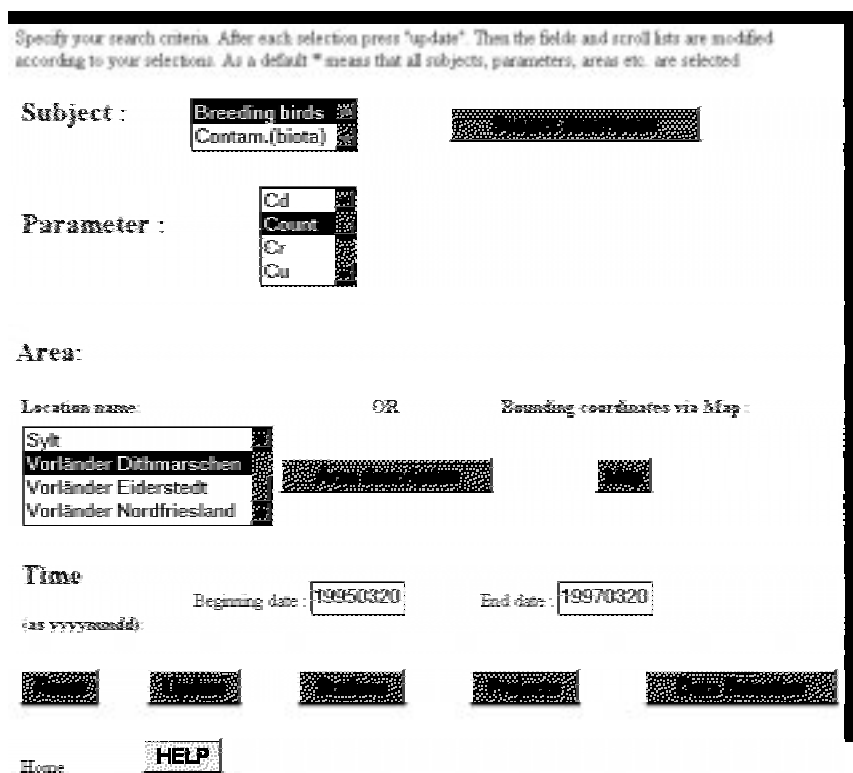
1. different structure of data (data models) at each location
2. different composition and processing state of data at each location
3. distribution of the relevant TMAP data at different trilateral sites

As a solution, the DEMOWAD project developed a data handling unit which can be implemented at each database comprising the following components:

1. The TMAP database (storage component); TMAP databases have identical, harmonized structures which map the trilateral exchange formats.
2. The 'integrated data catalogue' (catalog and retrieval component); it allows an overview about the availability, composition and processing state of the data and retrieval of selected data. The information for the data catalogue is stored in meta databases which are integral parts of the TMAP data bases.
3. Networking via Internet services (network component); it allows data transfer and interactive data catalogue access between the different TMAP databases and trilateral users.

This means, that at each national Wadden Sea database center, an extra TMAP database has to be maintained, which gets its data directly from the national database. The trilateral data exchange format is the direct consequence of the TMAP database structure. On top of the integrated data catalogue an interactive and dynamic World Wide Web (WWW) user surface exists (see figure 1), which guides the users, with the help of the catalogue data to the wanted parameter or value data. Internet services also enable the data transfer (data in user-adapted formats) to the user sites.

Fig. 1: Main search page of the DEMOWAD data handling unit.



The integrated data catalogue with the WWW surface guides the user interactively towards the wanted data. The main search topics of the data catalogue are:

- subjects (water chemistry, contaminants in biota, etc)
- parameters of a subject (e.g., heavy metals in mussels)
- location (areas or coordinates)
- time period

The elements of the menus always reflect the current content of the data catalogue (dynamic system) and the menus show only available information related to the selected search pattern.

E.g., if the subject 'contaminants in biota' in the area 'Danish Wadden Sea' is selected, only these parameters occur in the parameter selection menu, which are available for this area and subject.

The users will receive requested TMAP parameter (value) data together with meta data (information on data) in the common ASCII and dBase formats. Highly structured formats like the international standards ICES or HDF (Hierarchical Data Format) have to be developed in future. The data catalogue and data exchange format of the DEMOWAD data handling unit was developed in consideration of international standards to keep the unit open for other systems. This integrative approach enables other meta data catalogues in the three countries and the European Union to use also the DEMOWAD data catalogue for investigations.

The DEMOWAD data handling unit is based on a cost efficient solution. The results of trilateral user and environmental data center inventories, the inclusion of international data handling standards and the development in accordance with modern soft- and hardware technology characterize the trilateral data handling system and its advantages as follows:

- FLEXIBILITY:

The trilateral data handling system is decentral structured and is not limited to a fixed number of DEMOWAD data handling unit, although too many units involved will make the data search more awkward. Each database location able to implement the unit can join the trilateral data exchange. Modern technology and high modularity will keep the system flexible and open for further development.

- MINIMAL EXTENSION COSTS:

The data catalogue, networking, TMAP database structure and the principle of the data extraction have already been developed by the DEMOWAD. As a consequence of harmonized, subject oriented TMAP data base structures, the implementation of further parameter groups can be done once and straightforward for all implemented units.

- MINIMAL MAINTENANCE COSTS:

An automatically running updating procedure to fill in national TMAP data into the trilateral databases (restructuring national data) has to be developed at each national database once. The operation of the TMAP database itself needs only reduced effort, because in contrast of full working databases writing operations will occur seldom and are controlled exclusively by the national data centers.

- HIGH DATA SECURITY:

The TMAP data bases will use the high standards of existing data base security. From outside, the data cannot be changed. Efficient authorization and authentication procedures will be implemented.

- COMMON DATA FORMATS:

The TMAP data can be exported to the trilateral users for commonly used text processor and spreadsheet programs.

- CONNECTION WITH OTHER INFORMATION SYSTEMS:

The data catalogue system can be linked to other national, European or international data information systems (meta data only).

- USER-FRIENDLINESS:

The easily to be used WWW surface of the integrated catalogue database allows a direct and complete view into the available data and thus creates a high user acceptance.

## II. CLASSIFICATION OF A-PARAMETER DATA

In the prototype test, the indicated components of the DEMOWAD data handling units were specified and installed at the participating databases.

The basic units of the TMAP database are subjects, which are, in view of data handling, related parameter or parameter groups. This kind of model is easy to implement and easy to extend, and it is also simple to connect data with the catalogue system. This approach also allows the implementation of generic data structures, which can hold similar data of different monitoring parameters.

For the prototype test two subjects were implemented: '*breeding birds*' and '*contaminants in biota*'. For the last subject only data of the priority A-parameter 'heavy metals in mussels' was implemented.

## ANNEX 7

The experiences and evaluation of the prototype test revealed that the effort for further implementation of TMAP data depends on the general complexity and diversity of the data itself and if raw or aggregated data is wanted.

### Complexity and diversity of the A-parameter data

With the following data handling criteria the A-Parameter were divided up into five data classes. This classification simplifies an estimation of implementation efforts and allows an independent basic assessment of the A-parameter data.

#### A. Complexity of data structure

The complexity of the A-parameter structure increases with the number of co-variables and with the complexity of the measured parameter itself. Therefore, A-parameters with a lot of co-variables requires more effort than simple structured parameters.

Implementation effort: the more co-variables → the higher the data complexity → the more effort

#### B. Raw data or aggregated data

Raw data is the direct output from measurements (here: including quality and validity checks). It is always possible to make a new assessment of this data, if new scientific knowledge demands another analysis and calculation of this data.

Aggregated data is processed and summarized information of parameters including co-variables on a certain scientific background. Therefore, aggregated data has often a simpler structure, a smaller amount than raw data and is easier to be implemented. On the other hand, if other analysis demands different aggregation of raw data considerable effort will be needed.

Implementation effort: the more raw data → the more effort

#### C. Diversity of data structures

If different A-parameters have a similar data structure, the development of more generic data structures is possible. E.g., the data structure of the A-Parameter '*contaminants in mussels*' and '*contaminants in bird eggs*' can be summarized in the more generic data structure '*contaminants in biota*' and needs therefore less effort.

Implementation effort: the higher the diversity of data structures → the more effort

### A-parameter data classes

Four classes of A-Parameter were identified due to their criteria:

- A. complexity,
- B. expected storage status (raw or aggregated) and
- C. diversity of data structures.

#### Class 1: Biological Data

- A. complex structured data
- B. stored as raw data

One or more biological parameters in this class have a complex structure in the TMAP database, because for each parameter a wide variety of co-variables (habitat data, tissue data, etc.) has to be measured. The biological parameters differ in a wide range (e.g., data on migratory birds and data on phytoplankton) so that the data structures of the A-Parameter in this class is very heterogeneously.

Communities on tidal flats  
 Mytilus stocks  
 Fish key species  
 Phytoplankton  
 Zooplankton  
 Breeding birds/breeding success  
 Migratory birds  
 Beached bird survey  
 Seals/breeding success

**Class 2: Chemical Data**

- A. medium complex structured data
- B. stored as raw data
- C. homogeneous data structure in this class

Chemical data has generally less co-variables than biological data so that the TMAP database structure of this group is medium complex. Because these parameters always deal with contaminants in inorganic or organic matter, the data structures in this group is relatively homogeneous. The storage of the raw data is proposed.

- TBT in water
- Inorganic nutrients in water
- TBT in sediment
- Metals in sediment
- Metals in Mytilus
- Organochlorines in Mytilus
- PAH in Mytilus
- TBT in Mytilus
- Metals in fish
- Organochlorines in fish
- PAH in fish
- Contaminants in eggs
- Contaminants in seals

**Class 3: Habitat Data**

- A. simple structured data
- B. stored mostly as aggregated data
- C. homogeneous data structure in this class

Habitat data in the TMAP will mostly be considered as a course scale and therefore be included in the trilateral data handling system after preprocessing in GIS. The processed data will have only few co-variables and therefore exhibit a simple structure. For some parameter groups (e.g. salt marshes, dunes and geomorphology) not only table data is required but coordinates too (e.g. location and morphology).

- Geomorphology
- Salt marshes
- Dunes
- Blue mussels
- Sabellaria reefs
- Eelgras
- Macroalgae

**Class 4: General Data**

- A. simple structured data
  - B. stored as aggregated data
  - C. heterogeneous data structure in this class
- The group of general data are mainly used as additional variables for the assessment of parameters in the other data classes. The data of this class is not directly recorded for the Wadden Sea monitoring and is in most cases available as aggregated data on different stages. The wide variety of these parameters causes heterogeneous data structures.

- Hydrology
- Weather conditions
- Flooding
- Groundwater level
- Land use (types of utilization)
- Coastal protection measures
- Agricultural utilization (grazing of salt marshes)
- General fishery
- Human activities
- Flat walkers
- Ships in marinas
- Air traffic

**Coordinates**

The effort to store and process coordinates is very high because for further assessment GIS system with special databases analyzing and presentation tool are needed which require high expertise. Since in all Wadden Sea data centers the same GIS software (Arc/Info) is used, the structure of coordinates is basically homogeneous. Large implementation effort in each country is needed to harmonize different spatial scales, coordinate systems and data quality and to connect the GIS systems with the DEMOWAD data handling unit.



### III. STATUS AND AVAILABILITY OF TMAP DATA ON THE NATIONAL LEVEL

A national inventory of the available A-parameter data was made to get a first estimation of expected efforts to make TMAP data available on the national level. The aim of the data inventory was to get information on the physical storage status (where and how stored), the necessary steps to make this data available on the trilateral level and to get a rough estimation of the anticipated implementation cost. To simplify the evaluation of the inventory, the A-parameter data was classified according to their origin, structure and complexity in biological, chemical, geographical and general data.

Overview of the estimated efforts to make TMAP data (A-parameter) available on national level and to transfer national TMAP data into the trilateral TMAP database.

		high costs expected medium costs	high medium	low costs expected no information available	low	
		implementation effort				
A-parameter	Country	DK	NL	D/Nds.	D/S-H*	
<b>1. Biological data</b>						
Communities on tidal flats		high	medium	high		high efforts estimated
Mytilus stocks		medium	high	high		
Fish key species		medium	high	high		
Phytoplankton		high	low	high		
Zooplankton		high	high	high		
Breeding birds/breeding success		low	high	low		
Migratory birds		low	high			
Beached bird survey			medium			
Seals/breeding success		high	medium	medium		
<b>2. Chemical data</b>						
TBT in water			low	high		high efforts estimated
Inorganic nutrients in water		high	low	high		
TBT in sediment				high		
Metals in sediment		medium	medium	high		
Metals in Mytilus		low	medium	high		
Organochlorines in Mytilus		medium	medium	high		
PAH in Mytilus		medium	medium	high		
TBT in Mytilus			medium	high		
Metals in fish		medium	medium	high		
Organochlorines in fish			medium	high		
PAH in fish		medium	medium	high		
Contaminants in eggs			high	high		
Contaminants in seals			high	high		
<b>3. Habitat data</b>						
Geomorphology			low	high		high efforts estimated
Salt marshes		high		medium		
Dunes		medium	high	medium		
Blue mussels		medium	high	medium		
Sabellaria reefs		medium	high	high		
Eelgrass		medium	high	medium		
Macroalgae		medium	high	medium		

**4. General data**

Hydrology		medium		high efforts estimated
Weather conditions		medium		
Flooding	high	medium		
Groundwater level		high		
Land use (types of utilization)		high	medium	
Coastal protection measures	medium	medium		
Agricultural utilization (grazing salt marshes)	medium	high		
General fishery	high	high	medium	
Human activities	medium	high		
Flat walkers		high		
Ships in marinas		high		
Air traffic		high		

\* In D/Schleswig-Holstein a more detailed inventory was not available. A general assessment of the current state of data handling pointed out, that always high efforts are needed to make data available for the TMAP databases.

Only a rough estimation of the expected efforts to make data available at the national TMAP databases could be made, because the data is widely distributed and the structure and the completeness of the data is not known in detail.

1. For the most A-parameter concerning the biological data class, a high national effort is expected to make this data available for the TMAP databases. The reasons are the missing complex data structures of this data class and the respective programs which transfer the national data into the TMAP database and the unorganized storage status of the data.  
The general effort for biological data which are already collected in trilateral programs (birds, seals) is low, because of already existing data structures, data storage and data handling.
2. The effort to make chemical data available is comparably low, because this data is mostly well organized, stored in national databases. The data structures are, in contrast to biological data, less complex and programs which transfer the data into the TMAP databases are easier to be developed.
3. Habitat data is mostly stored in GIS (Geographical Information Systems) and therefore well organized. Because aggregated data is proposed, this data has to be calculated and transferred into the TMAP databases. The effort to make this data available is generally estimated as low, because GIS and the respective experts are available.
4. For the general data, the expected effort is estimated as low or high. The data is widely distributed between very different originators and institutions which in most cases do not corresponding with Wadden Sea interests. This is also the reason why for a lot of A-parameters of this data class information is not available.
5. The effort to include coordinates in the TMAP database is relatively high. It depends strongly on which data will be used. Therefore, the exchange of coordinates will be developed at a later stage.

The TMAP data inventory shows that the national data handling and data storage of TMAP data is insufficient to supply the requirements of the developed DEMOWAD Data Handling Unit. Therefore, the data handling of each A-parameter requires personnel and financial effort on the national level (see chapter VI of this annex). Because of already existing national Wadden Sea databases and the therefore better organized data handling, this effort is lower in Denmark and in The Netherlands than in Germany. The financial and personnel effort for the necessary national data handling could be reduced considerably, if the required databases, missing data structures and data organization would be developed and implemented in close cooperation with the trilateral data handling. E.g., trilaterally developed data structures, data exchange formats and database structures could also be used and implemented for national purposes.

#### IV. MAKING TMAP DATA AVAILABLE ON THE TRILATERAL LEVEL

With the results and experience of the prototype test the DEMOWAD data handling group identified four main technical tasks to make national A-parameter data available on trilateral level. An additional estimation of effort to implement a single A-parameter of each data class was carried out.

Ranking (from high to low effort ) and description of the tasks to make TMAP data available on trilateral level:

**1. Elaboration of the trilateral data structure for the TMAP database.**

Inventory of existing national and international data structures and standards. Consultation with experts of the respective A-parameter. Adaptation in the framework of the existing TMAP database. Alignment to the catalogue part of the TMAP database.

The development and implementation of trilateral data structures demands the most effort.

**2. Modifying the WWW/database program to allow retrieval of data.**

The program responsible for the user surface and the data retrieval has to be extended with the appropriate features to retrieve data of the new implemented A-parameter.

**3. Development of an interface program that transfer the national data into the TMAP database.**

Investigation of national database structures and development of programs and tools to convert and transfer the A-parameter data to the TMAP database (updating procedures).

**4. Design of the appropriate trilateral exchange format.**

The trilateral exchange format has to reproduce the TMAP database structure of an A-parameter in an appropriate data file format (e.g., ASCII, dBase).

For each data class a rough estimation of implementation effort was made. As criteria the complexity, effort on aggregation level and diversity of structures within one data class and the implementation results and experiences of the prototype test were used.

The effort was scored from **5** (high effort) to **1** (low effort).

Estimation effort of A-Parameter classes in relation to each other.

Data classes	Biological Data	Chemical Data	Habitat Data	General Data
Implementation Effort	<b>4</b>	<b>3</b>	<b>2</b>	<b>2</b>

For the implementation of the system, biological A-parameters need a lot of effort, because of their complexity and number of co-variables. Chemical data is also complex, but needs less effort, because the data is more homogeneous and can be implemented in generic structures. The geographical and general data class need low effort, because this data is simple structured and aggregated.

For the operating of the system, the effort is low in regarding the implementation of the trilateral data handling system.

## V. FURTHER IMPLEMENTATION OF THE TRILATERAL DATA HANDLING

Three steps are necessary for a successful trilateral data handling, based on the results of the DEMOWAD project and its prototype test.

1. The DEMOWAD data handling unit has to be installed at the Wadden Sea data centers.
2. The A-parameters have to be implemented.
3. The system has to be maintained and kept in operation.

Ad 1. The development of the DEMOWAD Data Handling Unit has been finished. The prototype test with the indicated two A-parameters proved the functionality, capability and reliability of the system. The units were installed at the national Wadden Sea databases in Denmark and in The Netherlands. One Unit was set up at a German database, but only for the DEMOWAD project. For the future Germany needs a decision on where to install the DEMOWAD data handling units and which national databases should be used.

Ad 2. From the beginning, the DEMOWAD units were kept open and flexible for further development and implementation of new TMAP parameters, which should take place as follows:

- Implementation of the two prototype A-parameters (already finished).
- Trilaterally harmonized A-parameters with trilateral guidelines (seals and migratory birds) will follow.
- The other A-parameter will be implemented closely related to the further TMAP.

It is proposed that data (mainly of the general data class), which is not security sensitive, low in number, highly aggregated and not related to the Wadden Sea databases and which installation at each data center has no cost benefit, should be collected, stored and maintained at one place.

If national Wadden Sea databases are not available (Germany) or if national data structures are not defined (for a lot of A-parameters), it is recommended to develop these databases and structures in close cooperation with the trilateral data handling. To reduce costs, the structure of the TMAP database or trilateral data structures could be used on the national level.

ad 3. After implementation of the A-parameters the trilateral data exchange system has to be maintained and regularly updated.

## VI. REQUIREMENTS AND EFFORTS OF A TRILATERAL DATA EXCHANGE

### National level

The DEMOWAD data handling project developed a trilateral data exchange system which solve the problem of the heterogeneous national Wadden Sea data storage. The presented data handling unit can be installed under the minimal technical and personnel prerequisites at each national Wadden Sea database location. The national databases and data centers remain independent and autonomous. Therefore, the organizational structure of the data handling in each country has not to be changed or modified to meet the needs of the trilateral data processing.

#### a. Technical requirements

The DEMOWAD data handling unit can be implemented at each national database where already validated and quality checked TMAP data has to be available in proper time. Technical requirements are:

- procedures to restructure national data into TMAP data and transfer it into the TMAP database for up-dating
- a server (computer) with appropriate operating system
- a relational database management system; in case of security policies the server has to work outside the firewall to allow Internet access
- Internet access
- programs for the Internet services

#### b. Administrative and personnel requirements

The DEMOWAD unit entails the technical prerequisites of a the trilateral data exchange in the TMAP. However, a running and an effective system is only possible if administrative agreements have to be reached between all different database locations at the national level on common procedures of data copyrights, data delivery, data processing and data assessment for the TMAP data. Experienced personnel is needed which guarantees a continuous operating and up-dating of the DEMOWAD unit and which organizes and develops the data transfer procedures between national and TMAP databases and trilateral working groups.

#### c. Implementation effort

The further implementation of the DEMOWAD unit could be done by projects or as part of the national database management.

Implementation costs depend on:

- number and complexity of parameters,
- expansion of system functionality (statistical, analytical and graphical tools, GIS),
- adoption of trilateral data structures into national databases (no effort on update procedure).

**d. Operation effort**

During operation only a relatively small effort is anticipated to execute the trilateral data exchange. The effort depends on the number of parameters and the frequency of up-dates of the catalogue and parameter data of the TMAP database (e.g. whether once or twice per year) and the amount of support required by the trilateral user groups.

**e. Financial and personnel consequences**

Costs for set-up of the basic data exchange system at the national databases:

- technical equipment (hardware and software for appropriate server),
- adaptation of the DEMOWAD unit at the national databases.

Cost during operation of the system:

- experienced database expert (1/2 person) who is responsible for the maintenance of the trilateral data handling tools, for regular up-dating of the TMAP data base, for the implementation of new parameter groups and for the expansion of system functionality,
- Internet access (e.g., permanent line to provider if not been available yet), additional data transfer costs.

### **Trilateral level**

To coordinate the operation and further development of the system,

1) a TMAP data handling group as a technical working group consisting of members from all participating data centers is needed to carry out the

- implementation of further parameter structures,
- development and implementation of further trilateral data exchange tools,
- maintenance of the trilateral data handling tools.

2) a full-time TMAP data handling coordinator is needed to organize the

- further development of the trilateral data handling tools,
- the implementation of further parameter groups,
- the continuous operation of the whole trilateral data exchange system,

and to support the trilateral working groups and the CWSS.