

# Exchange and Archiving of Measurement Data within the Research Project PROMORPH

By INGO JUNGE and BJÖRN SCHUBERT

## S u m m a r y

Coastal engineering research projects produce increasing amounts of data during field campaigns. In many cases data are collected and stored at different places. In the context of the research project PROMORPH – PROgnosis of medium-term coastal MORPHology, two different internet based data management systems were developed and tested at the Institute for Fluid Mechanics and Computer Applications in Civil Engineering of Hannover University. Both systems provide efficient data administration and sustainable archiving. In addition, a simple information and data exchange should be warranted between all project participants working at different places. Further functionalities such as searching and visualization tools were implemented to promote the general acceptance by users.

## Z u s a m m e n f a s s u n g

*Wasserbauliche Forschungsprojekte werden zunehmend durch umfangreiche Naturmessprogramme begleitet, deren Messdaten häufig an unterschiedlichen Orten gesammelt und vorgehalten werden. Im Rahmen des Forschungsprojekts PROMORPH – PROgnose mittelfristiger MORPHologieänderungen – wurden zu diesem Zweck am Institut für Strömungsmechanik und Elektronisches Rechnen im Bauwesen der Universität Hannover zwei internetbasierte Datenmanagementsysteme entwickelt und getestet. Neben einer effizienten Datenverwaltung und -archivierung wurde im Besonderen Wert auf einen reibungslosen Informations- und Datenaustausch zwischen den Projektteilnehmern gelegt. Darüber hinaus sollen weitere Funktionalitäten, wie z.B. Such- und Visualisierungsmethoden die allgemeine Akzeptanz der beiden Datenhaltungssysteme erhöhen.*

## K e y w o r d s

Data Base, Data Management, MetaView, Promorph

## C o n t e n t s

1. Introduction .....	340
2. Concepts of the PROMORPH Data Base System .....	341
2.1 Description of the Decentralized Data Management in PROMORPH .....	342
2.2 The Data Base Input Tool.....	343
3. The Data Base MetaView.....	345
3.1 Concepts of MetaView.....	345
3.2 The Input of Data Sets.....	345
3.3 Data Search and Download.....	347
3.4 Visualisation of Data .....	348
4. Conclusion.....	350
5. References .....	351

## 1. Introduction

Today's coastal research projects tend to handle continuously growing amounts of data, which require efficient data administration and archiving. At the same time, simple and reliable data exchange plays a substantial role for suppliers and/or data users who frequently work at distributed places. Against this background storage of project data with an access via the internet offers a practical solution for data management, availability to all project participants and for sustainable archiving.

With the design of a decentralized data management questions concerning the requirements and the function range generally arise: Is the system primarily intended for data exchange, or should further options be made available to the user, e.g. for data search or visualization of data or merely to display meta-information? In case of visualization one has to consider that input data probably require reformatting or preprocessing. This additional effort usually has to be carried out by the data supplier and would greatly reduce the acceptance of a system. It may even lead to a certain hesitance to update the data bank frequently. Therefore, a system has to be designed, which requires minimum effort for data input.

Information systems, which access externally stored data of associated data suppliers have great advantages. The recently established North Sea and Baltic Sea Coastal Information System NOKIS (LEHFELDT and HEIDMANN, 2003) connects independent meta-data bases generated and maintained at various institutions in the German coastal zone. Data can be accessed via the internet through a web portal <http://nokis.baw.de>. The data suppliers use efficient data bases locally for the administration of mass data, which prevents data duplication and improves acceptance. However, it remains difficult to design a balanced system, which offers both simple and fast data input to the data supplier and, at the same time, convenient methods of a high performance information system. A good example of the convenient range of functions offered by a data base for hydraulic engineering is the information system developed within the KFKI-project MORWIN (LEHFELDT and BARTHEL, 2001).

In the PROMORPH research project (ZIELKE, 2002), two separate internet based data management systems (DMS) were developed and tested at the Institute for Fluid Mechanics and Computer Applications in Civil Engineering of Hannover University, Germany. The project, funded by the German Federal Ministry of Education and Research (BMBF) started in 2000 and was successfully concluded in 2002. Because of the multitude of data collected within the project in extensive field campaigns and from other sources (ALR Husum, BSH), an efficient system for data administration and sustainable archiving had to be developed. In addition, a simple data exchange was to be ensured between all project participants working at different locations. A voluminous data pool forms the basis for numerical modelling of hydrodynamic and morphodynamic processes in the investigation area and the exchange of project data via internet is a basic feature. Therefore, a secure and limited access to the data pool has to be guaranteed by the data base system. The proposed system differs from other information systems (e.g. KLIWA project – Arbeitskreis KLIWA, 2001, Umweltdatenkatalog – Nouhuys, 1998), which only support users in searching for data but offer no download options. The PROMORPH data base is predominantly designed for data exchange and archiving within a small project group consisting of a manageable number of participants from different locations. During the PROMORPH project about 15 scientists worked at 5 locations. Considering the fact that all data suppliers are well known, quality assurance of data base content plays a minor role.

In the first phase, a concept is introduced which is based on a hierarchical administration of HTML pages guiding the user through the data description (meta-information) and offer-

ing access to the externally stored field data. A basic version of this concept was successfully implemented in the research project COSINUS (MARKOFSKY, 2000).

During the PROMORPH project further development of this system confirmed simple handling and a high flexibility concerning the meta-data input. However, flexibility of the input also leads to a non-uniform structure of meta-data, which causes increasing efforts related to data searching, also impaired by a lack of search functions. For this reason, a second DMS called MetaView ([www.metaview.de](http://www.metaview.de)) was developed. Here, the meta-information is stored in a relational data base (MYSQL) whereby convenient search functions can be implemented and offered to the user. MetaView is a long term application, where meta-information about collected and archived project data is made accessible to the general public via this data base. In addition, further data collections from other scientific projects can be stored and administered separately by the system.

In the following section 2, we discuss set-up and functionalities of the first data management system. Section 3 presents the MetaView data base system. The paper finishes in section 4 with notes on the current state of both system solutions.

## 2. Concepts of the PROMORPH Data Base System

It is important to distinguish between data and meta-data. Data are the values/numbers that were actually measured or calculated. Meta-data are the descriptions, in this instance stored in HTML text (Hyper Text Markup Language) and made accessible via the internet (see Figure 1). For the creation of these HTML documents, the data management system provides an editor as input tool to the data supplier, so that knowledge of HTML programming is not necessary.

For the structuring of meta-data the user has the option of organizing the text in hierarchically ramified levels. This opens up the opportunity of constructing a ramified description which helps to supply explanations relevant to a data group of a higher level and to supply special information for individual data sets of a lower level. In this way an extensive data description can be clearly structured into a report. At the end of a (sub-)chapter a link leads the user to the desired data stored on the (external) server of the respective data owner.

The decentralized concept of web-based data management was chosen for the PROMORPH project for the following reasons:

- The data are stored locally at their source which enhances the acceptance of such a system by the data owner as compared to a centralized solution. This also facilitates maintenance and up-dating of the data base.
- The visualization of research results and exchange via the internet are state-of-the-art methods and indispensable tools for scientists.
- Already existing HTML pages can be integrated into the data description by setting respective links. Thus, the data description can easily be extended.
- Before an authorized user can access data he is always led through the data description.
- Each PROMORPH project partner can read all the meta-data but can only change his own data.

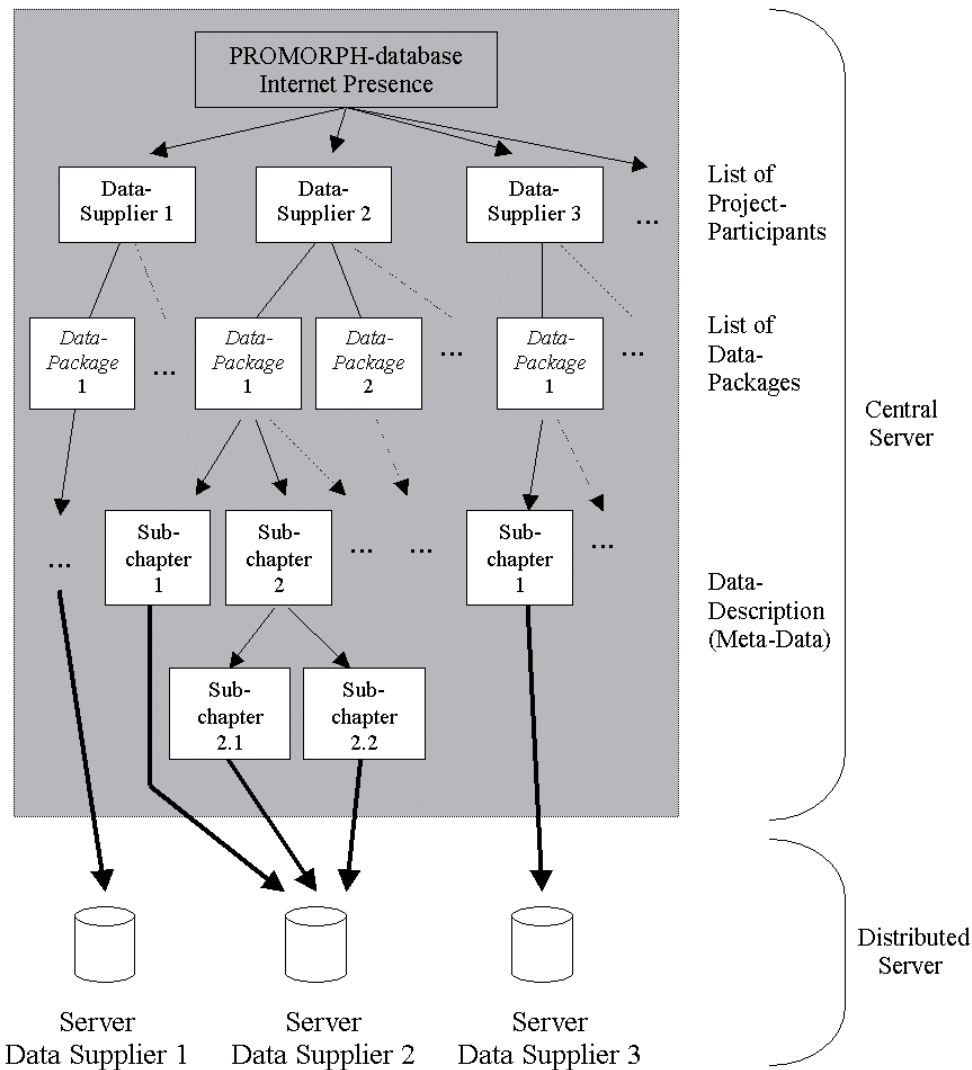


Fig. 1: PROMORPH Data Base Concept

## 2.1 Description of the Decentralized Data Management in PROMORPH

The data in the PROMORPH data base are organized in data packages. Each package consists of one or more hierarchically structured HTML pages which supply information (meta-data) on the current data.

All institutions involved in the project are listed in a table on the data base start page. After selecting a participant, a list of data packages supplied by the partner opens. Behind each package, the name of the person who created the package appears in parentheses.

Selecting a data package first leads to the data description (see Fig. 3). If the meta-data were subdivided into different sub-chapters, one can find the link to the sub-chapters at the bottom of the page and the link to the previous page at the top.

The sub-chapters are structured identically. At the end of a data description the user will find a link to the desired data. Usually a password query takes place. Before setting the data link, the user has the possibility of summarizing essential meta-data (e.g. measurement loca-

tion, time, instrumentation etc.) in a header table. It is not necessary to fill in such a table. It only clarifies and simplifies access to keywords with an optional implementation of search algorithms.

## 2.2 The Data Base Input Tool

The input of meta-data is supported by the input tool, a simple Java applet editor, which is available through the data base internet page. In the first step, an access window opens and checks the access authorization of the user by a password query. This procedure assures that a data base user can only edit his own packages (meta-data and links to the current data). When the login is successful, an input window appears in which the user can enter and modify his meta-data (see Fig. 2). The input tool offers options to:

- create a new package
- create a new sub-chapter
- delete a chapter (with all sub-chapters)
- edit text into a chapter
- connect an external data file to a string in this chapter
- connect a reference to a string in this chapter
- rename chapters
- generate the status last stored (undo function)
- create a header-table
- select meta-information from data files.

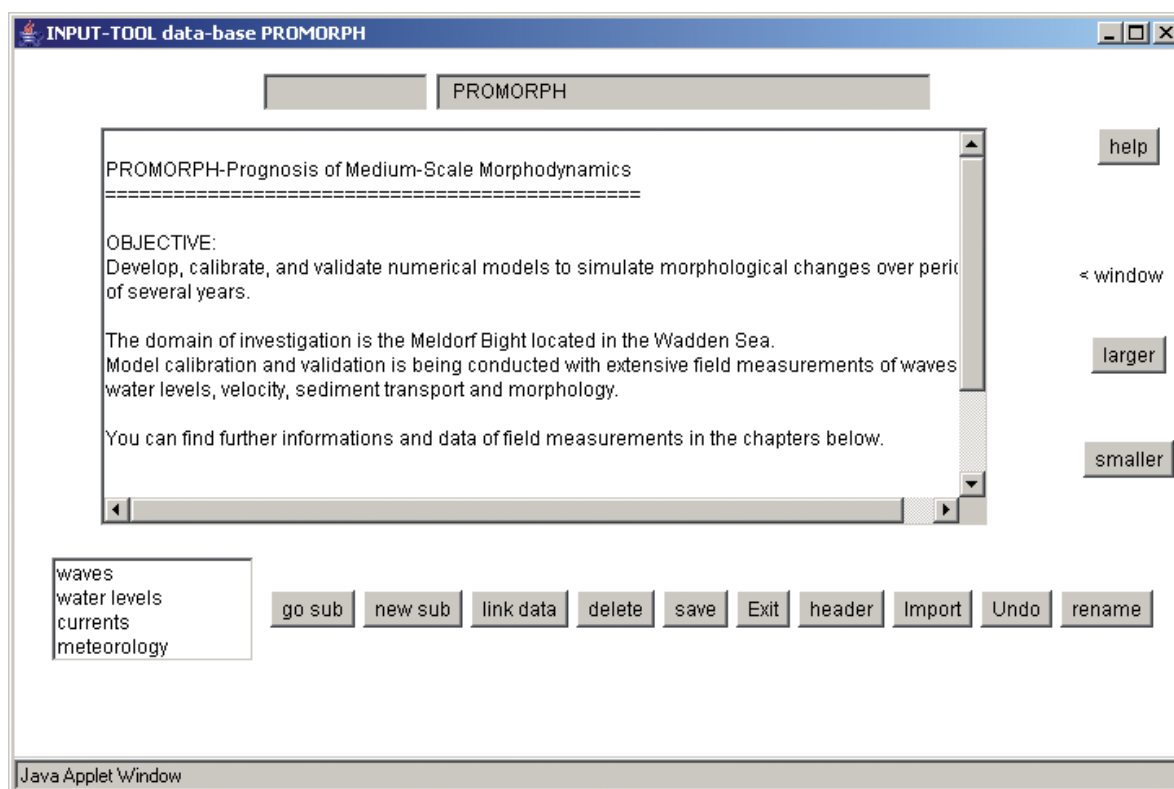


Fig. 2: Entry of Meta-Data via the Input Tool

The header table outlined above is useful to summarize the essential meta-data in a compact table. The input tool offers a mask with 24 possible entries to be filled in by the user.

Another helpful function of the input tool is the possibility to extract meta-information from a set of data files, which show a uniform format. In many cases, data files contain introductory information referring to the measurements (time, place, instrumentation, data-owner etc.). This information can be selected by using a function called import tool. This tool searches in pre-selected data-files for previously defined keywords. When a keyword is detected, the associated line with meta-information is transferred into the input window. It is also possible to extract a predefined number of lines from each data file.

When all meta-data and required links to the data proper are entered, the input must be saved. The input tool then converts the chapter into a HTML document, which is shown after each storage behind the input window in a second frame (see Fig. 3). This allows the user to verify the completed HTML-page and modify the result if necessary.

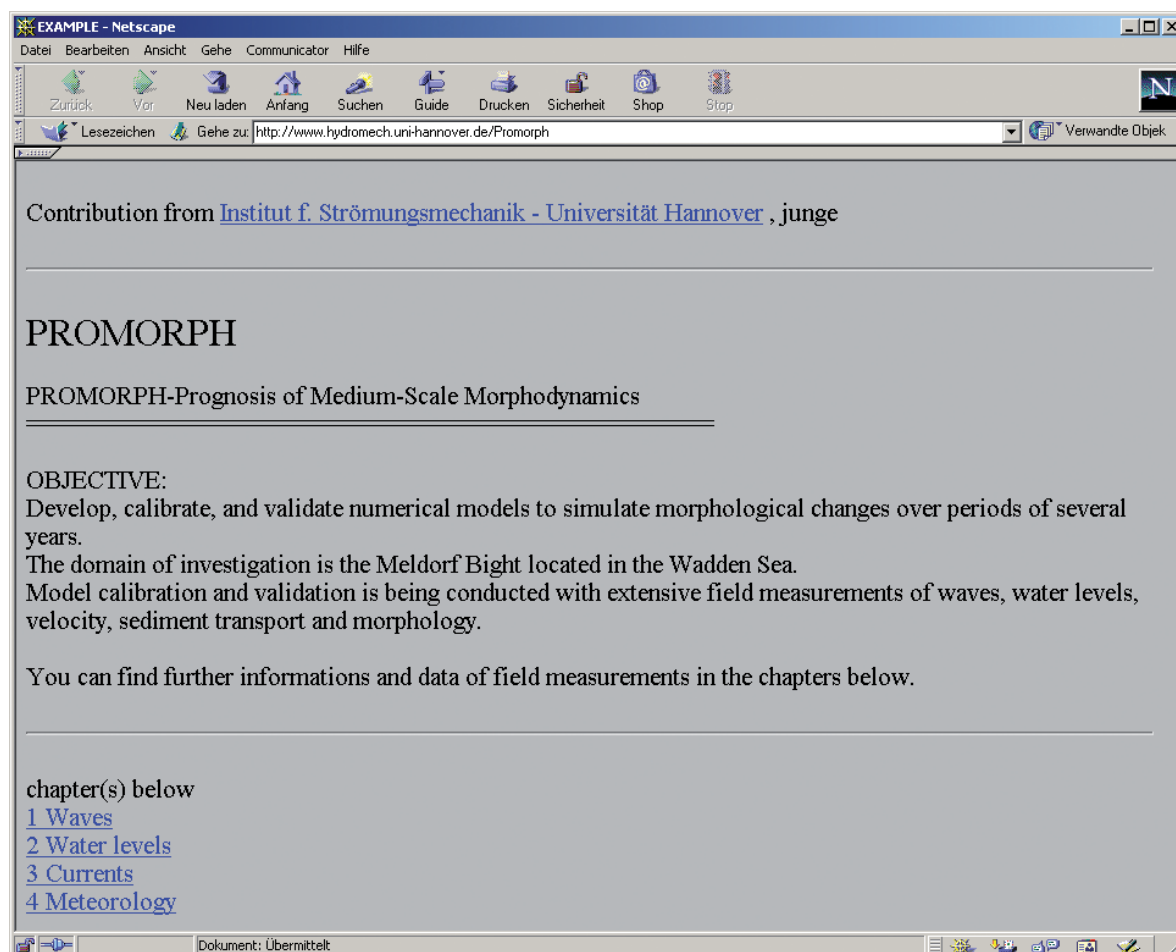


Fig. 3: Meta-Information in the PROMORPH Data Base

### 3. The Data Base MetaView

The data management system introduced in the previous section has many advantages. The simple and unformatted input of meta-data permits a fast but non-standard adapted input of meta-information. However, this flexible input leads to a non-uniform structure of meta-data which implies an increasing orientation effort for the data search, impaired by the non-existence of search functions. Due to the present amount of available data, this aspect is not significant for the PROMORPH project participants. In consideration of the fact that project data will be made accessible to the general public, it is vital to ensure an easy navigation within the data description, also to non-project members. For this reason a second internet-based information system named MetaView was implemented (see Fig. 5). It has access to the same external data pool but provides the user with improved and more extensive functionalities.

#### 3.1 Concepts of MetaView

The MetaView system only contains meta-information about measurement data stored inside a relational data base system. The original measurement data sets remain on the data servers of the participating institutions. Thus, necessary hardware resources have been distributed between several institutions. The structure of the system is illustrated in Fig. 4. It largely resembles the described PROMORPH data base. The significant difference is storage of meta-information in a MYSQL data base instead of using unstructured hierarchically linked HTML pages.

In order to generate the meta-information from measurement data sets, the MetaView system offers an integrated import module, which converts any measurement data file into a uniform format. As an important prerequisite of the system a file format called VA format was developed at the Institute for Fluid Mechanics and Computer Applications in Civil Engineering of Hannover University. A detailed description of the VA data format is given by ZIELKE (2001) and the online system information at [www.metaview.de](http://www.metaview.de). By converting all data sets registered in the MetaView system into one common format, the user is able to access data sets from different institutions in the same format. In addition to the download facility, the MetaView system offers a graphic tool for quick visualization of selected data sets.

#### 3.2 The Input of Data Sets

When adding new data sets to the MetaView system, a data supplier has first to generate a new group of data sets and attach meta-information concerning the measurement, i.e. time, location, data type and range. If the system already contains files of similar format, the data supplier can select an existing import mask, otherwise the design of a new mask is necessary. Finally, the user must specify the URL address to the external data server and the password if the access is restricted. If the data directory of an external server contains further files in addition to the desired ones (e.g. protocol files), the usage of so-called regular expressions allows access to a specific selection of data files. In this case, the MetaView system considers only those file names, which contain e.g. a given sequence of characters.

A further helpful parameter is the so-called scan cycle which can be defined for each

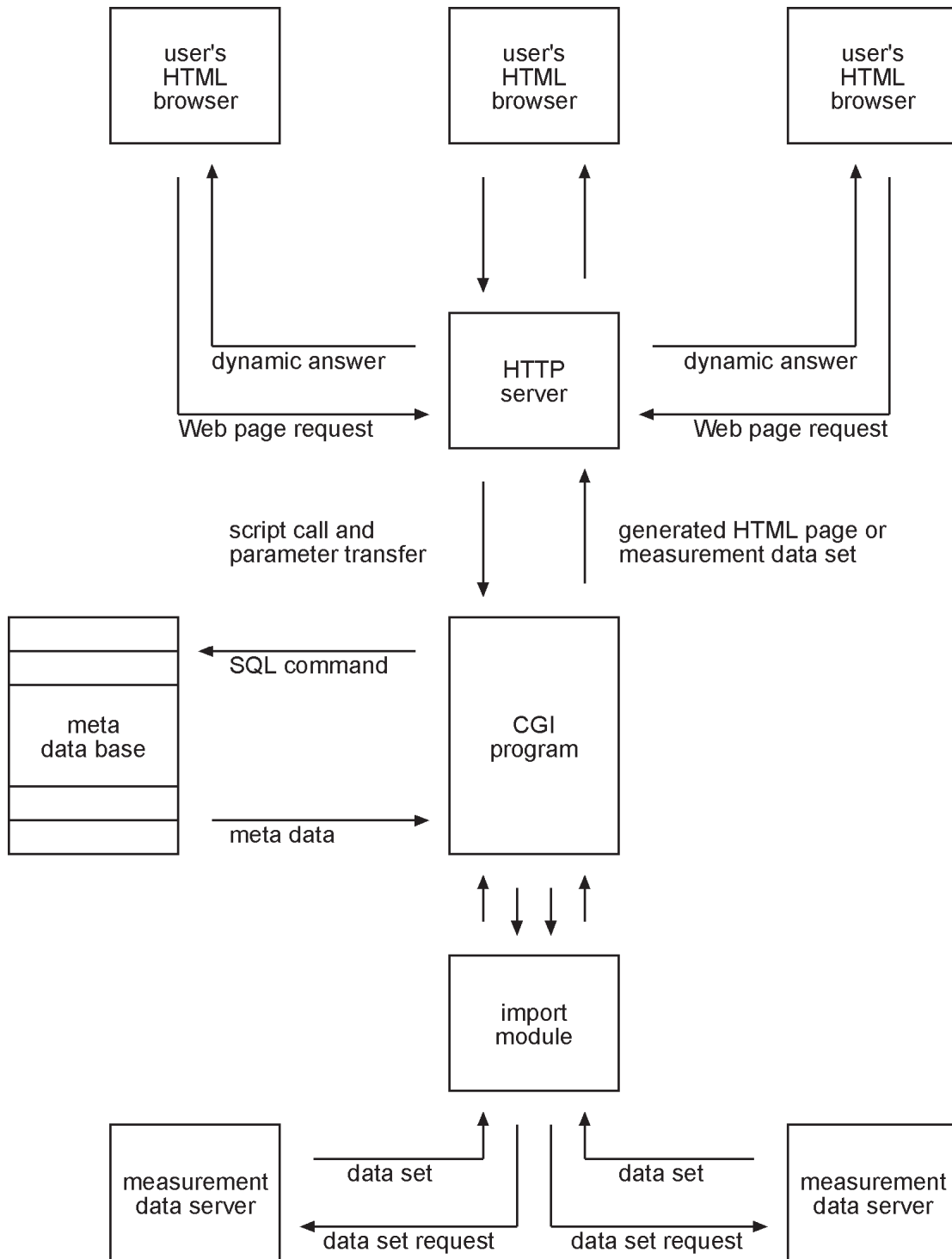


Fig. 4: Structure of the MetaView system

group of data sets. It schedules the dates when the system has to search through the directories of external data servers for new measurement data sets. The MetaView system is able to automatically read the new data records with the help of an appropriate import mask. This method should help to minimize the input effort for further measured data. Continuously recorded data (e.g. weather station) are added to the appropriate directory on the data server. MetaView checks these directories in given time intervals and reads new entries independ-



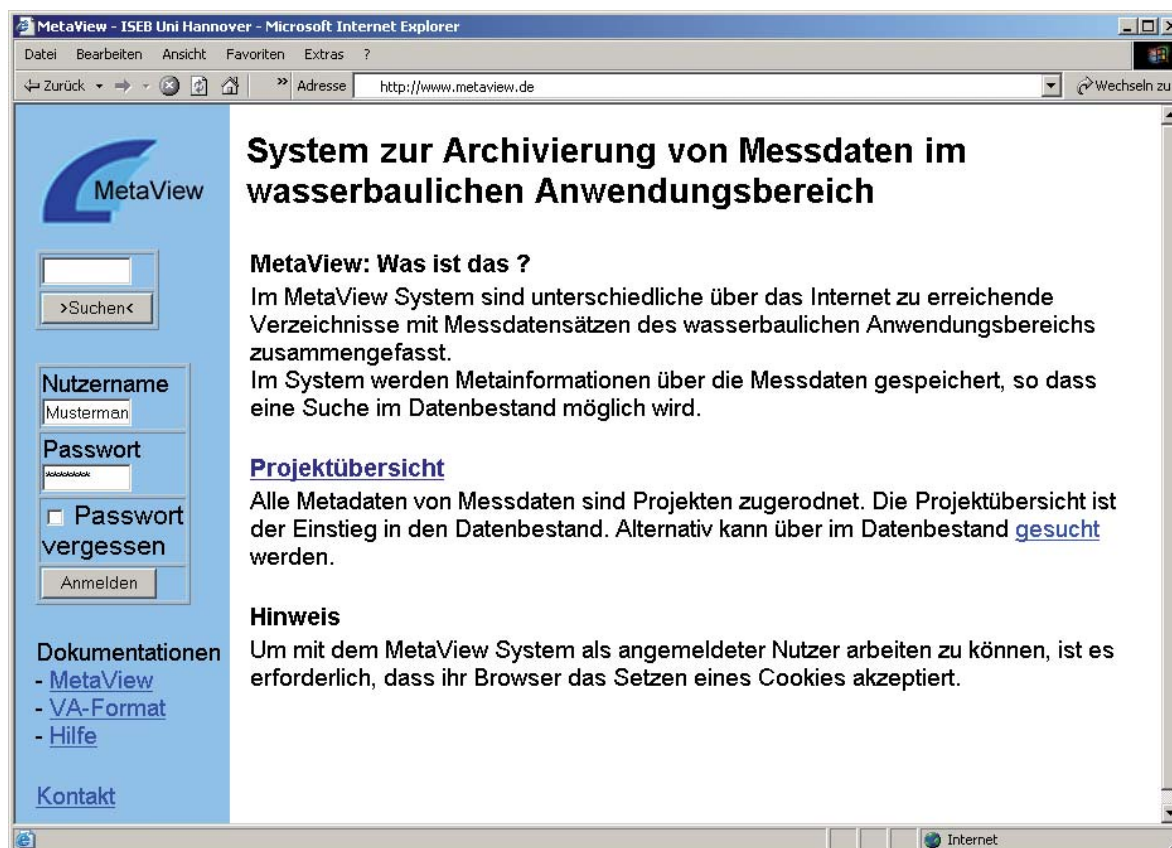


Fig. 5: Initial MetaView window

ently, if desired. If the data files contain special information concerning the measurement (e.g. time, location, etc.), the system can extract and process this meta-information with the help of the known import mask. Thus, the content of MetaView grows automatically.

Access to measured data can be restricted by an integrated user administration and access control system in order to make measurement data available to selected groups of users. The user administration allows to define single users and also groups. Thus, only little effort is required to define access to data sets for different users. During the project period, access to this data collection is restricted to the project participants. Later on, access can be granted to the general public by a simple adaptation of the access rights.

### 3.3 Data Search and Download

All measurement data to be extracted via the MetaView system are assigned to individual projects. Therefore, it is possible to collect further hydraulic data within the scope of similar research projects or other suitable measuring campaigns and to administer them by the MetaView system independently from the PROMORPH data pool. Each entry in the project list contains a description of the project and a list of groups of data sets already mentioned above. A group of data sets consists of a number of data sets, which represent the currently available measurement data files (see Fig. 6). In practice, several uniform data files pertaining to a measuring campaign (ADCP measurements, water level, sea state, etc.) are

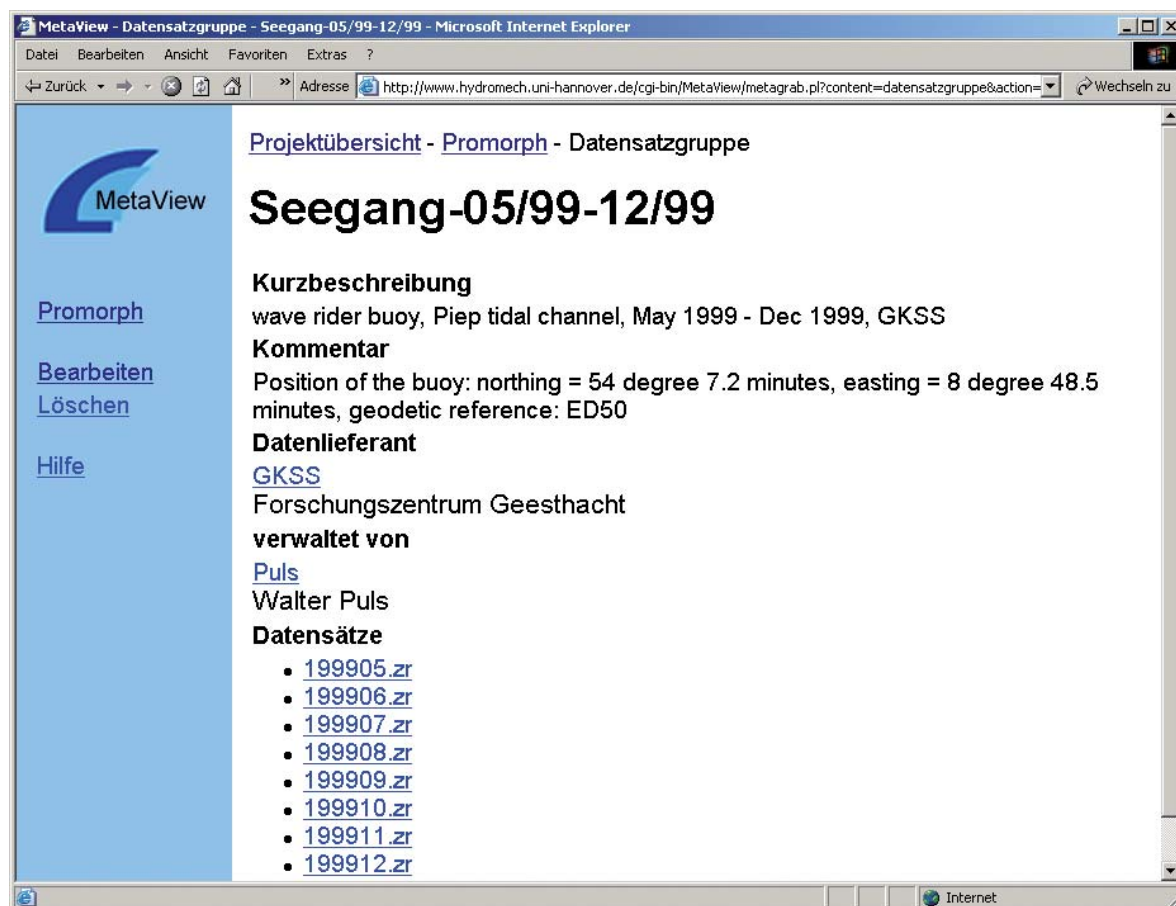


Fig. 6: Group of data sets containing sea state measurements over eight months

generally stored in a separate server directory. In the MetaView system, a group of data sets corresponds to such a directory, where each data set represents one individual measurement data file.

For an efficient data search, a user only needs to examine the project list and the list of available groups of data sets. This flat hierarchical structure allows fast access to measurement data. Alternatively, the user can utilise a convenient search function, which allows a full text search or a restricted search for projects, group of data sets, data sets, or also data suppliers. The user receives data files in the original data format or in a converted VA format. Each download is logged by the system in order to ensure an overview of the use of the measurement data sets. For this purpose, the user and the selected data format (VA or original) are stored together with the current time.

### 3.4 Visualisation of Data

The internal conversion of all data files into the uniform VA format facilitates visualisation of data. For this purpose, the MetaView system offers a special Java applet. In the simplest two-dimensional case, this tool selects the converted measured data suited for graphical presentation and lists all different measurement categories in two equal tables. The user can select a measurement category from each table, which defines the two axes of a simple xy-diagram (see Fig. 7), in the presented case a time series of wind velocities. In addition, the

choice of more than one variable in the second table is also possible and permits plotting of several curves in one diagram. By a repeated selection of other measured variables in the table beneath the plot, a quick and convenient visualisation of further measurement categories is possible.

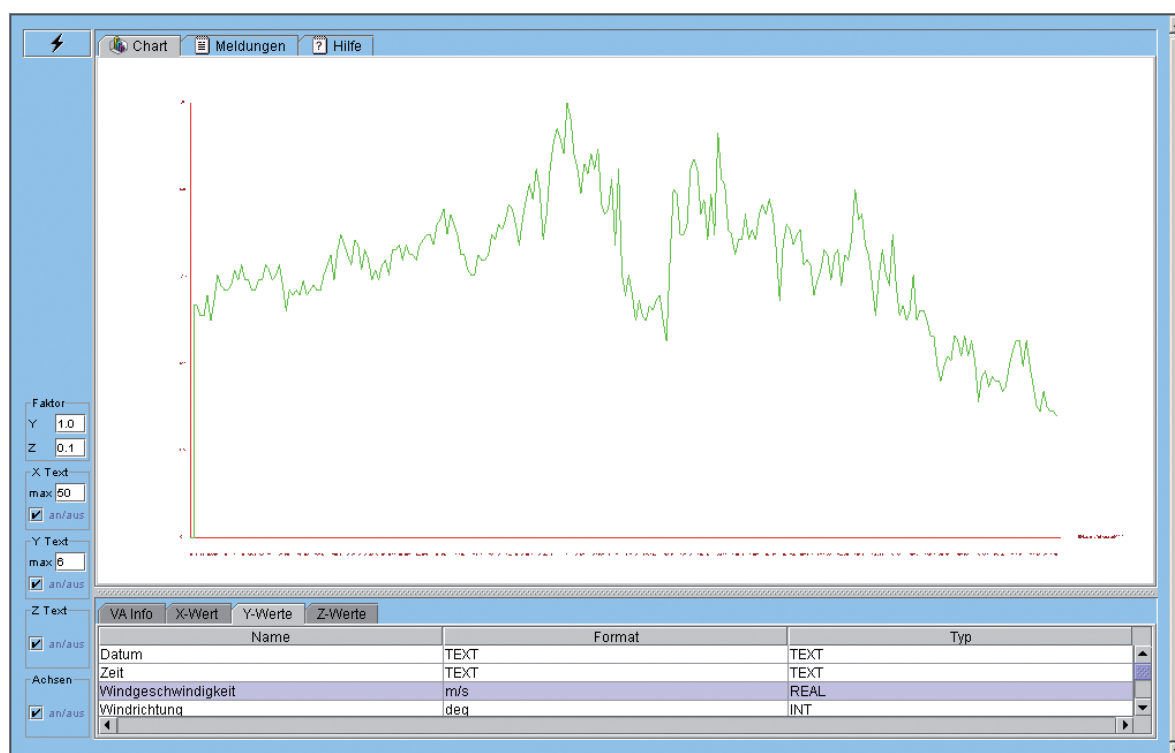


Fig. 7: Presentation of a simple xy-plot in the MetaView system.  
This example is a 24-hour distribution of wind velocity at station Büssum

Three-dimensional data can also be visualised under certain conditions. Many measurement data files may consist of a sequence of several data blocks. In this case, a data block generally contains the results of measurements at one location with each subsequent block valid for a different time window or a different location. The graphic tool permits representation of two measured variables of one data block as an xy-plot. A third variable distinguishes the individual data blocks (e.g. time or local specification) and is then selected for defining a third axis. As an example, Fig. 8 shows continuous vertical velocity profiles in a cross-section of the Norderpiep as a result of an ADCP measurement. Each individually measured vertical profile corresponds to an individual data block. The user can vary the number of displayed verticals by selecting a respective number of data blocks from the table below. A display of individual data blocks is also possible. Conveniently, the plot can be rotated, shifted, and scaled by mouse control.

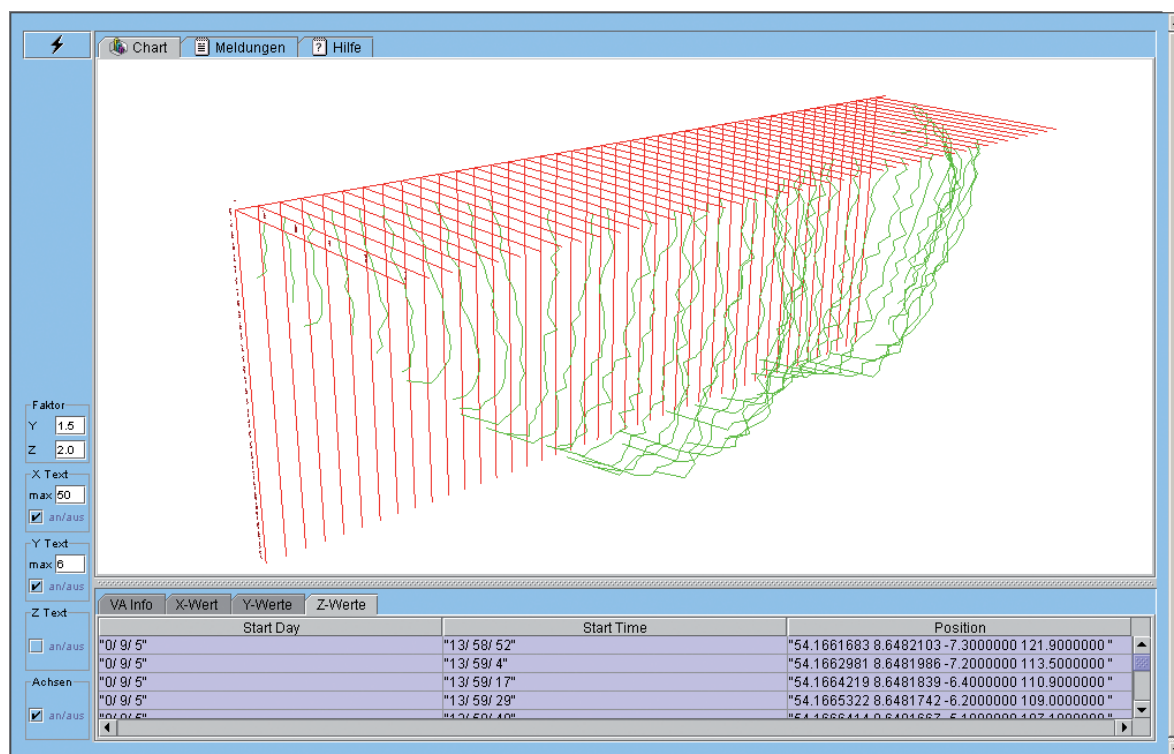


Fig. 8: Velocity distribution in a cross section of the Norderpiep in the Meldorf Bight. The visualisation tool for vertical velocity profiles provides an efficient data quality check

#### 4. Conclusion

The current data volume of both data management systems includes ADCP measurements, waves, meteorological conditions, bathymetric surveys and results from CTD measurements. More than 5400 data files can be accessed by the participants of the PROMORPH project. The measurement data can partly be viewed (only MetaView) and, if desired, downloaded by the user. The MetaView system provides access to the same measurement data pool as the first PROMORPH data base introduced in section 2. Project data are externally stored on servers at participating institutions. A redundancy of measuring data due to the two different data management systems does not appear.

It can be expected that the content of the MetaView system will increase continuously due to future measurement campaigns and the possible listing of other research projects. Presently, the concept of the MetaView system is used to set up a data base to support a wind energy research project also funded by the German Federal Ministry of Education and Research (BMBF).

## 5. References

- ARBEITSKREIS KLIWA: Meta-Informationssystem für KLIWA-Daten, Forschungsvorhaben KLIWA-Klimaveränderung und Konsequenzen für die Wasserwirtschaft, Bereich C, Projekt C4, Statusbericht 2001; <http://www.kliwa.de/de/>.
- LEHFELDT, R. and BARTHEL, V.: MORWIN – Morphodynamische Modellierung von Windwat-ten – Netzgestützte Projektbearbeitung im Küsteningenieurwesen. Abschlussbericht 1997–2000, 2001.
- LEHFELDT, R. and HEIDMANN, C.: NOKIS – Erstellung eines Metadateninformationssystems für die Küstenforschung und das Küsteningenieurwesen. Zwischenbericht 2003, 2003.
- MARKOFKY, M.: The COSINUS Database for Cohesive Sediment Transport in Estuaries and Coastal Zones. 6th. Int. Conf. On Nearshore and Estuarine Cohesive Sediment Transport INTERCOH 2000, Delft, September 2000.
- NOUHUYS, J. et al.: UMWELT-Datenkatalog des Bundes, FuE-Vorhaben 297 12 120/03; Endbe-richt, 1998.
- SCHUBERT, B.: Entwicklung eines internetbasierten Metainformationssystems mit dezentraler Datenhaltung für den wasserbaulichen Anwendungsbereich. Master Thesis – Institute for Fluid Mechanics and Computer Applications in Civil Engineering, 2001.
- ZIELKE, W. et al.: Datenmanagementsystem zur Sicherung der Erfassung, Verfügbarkeit und Vergleichbarkeit von Messungen in kontaminierten Grundwasserleitern. VEGAS-Ab-schlußbericht 1999, Institut für Strömungsmechanik, Universität Hannover, 1999.
- ZIELKE, W. et al.: PROMORPH – Prognosis of Medium-Term Coastal Morphology. Final Re-port, Institute for Fluid Mechanics and Computer Applications in Civil Engineering 2003.