PANGAEA

Publishing Network
for Geoscientific and Environmental Data

System Description and User Manual

This document is a working draft!

Alfred Wegener Institute for Polar and Marine Research (AWI)
27515 Bremerhaven, GERMANY

Center for Marine Environmental Sciences (MARUM)
University Bremen, 28359 Bremen, GERMANY
1 SYSTEM DESCRIPTION

1.1 Introduction

The information system Pangaea is a digital library with open access to archive, distribute and publish geo-referenced data from basic research in earth and environmental sciences.

In 1993 the research institutes AWI and MARUM established a data archive for general use by the scientific community, initially funded by the German Ministry for Research and Technology (BMBF). The system is based on a flexible and strictly normalized data model to fulfill the requirements of a heterogeneous and dynamic world of scientific data. Easy access to consistently formatted data is realized through a search engine in combination with a thesaurus accessing a relational database. The definition of individual configured sets of data is enabled through a data mining tool.

The system is fully operational since 1997. A web-server enabled access through the Internet via the domain ‘Pangaea’ (and ‘pangea’) in 1998. Since then the system is used by national and international projects and individual scientists to make data public available. In 2000 the World Data Center for Marine Environmental Sciences (WDC-MARE) was founded as part of the WDC-system of ICSU (International Council of Science). WDC-MARE uses the information system Pangaea as its central archive. The long-term operation of WDC-MARE and Pangaea is ensured through a commitment of AWI and MARUM.

The Pangaea data model includes all fields and options necessary for a complete data documentation which is crucial for the understanding of the data by the users of the library. Those metadata include at a minimum the geographical coordinates of the site of investigation, the principle investigator (PI) and a definition of the parameters with unit. References, methods, quality flags, date/time, elevation and comments/descriptions are optional to complete the metadescription. If data are unpublished or under moratorium, it can be password protected.

To ensure a reliable long-term availability of its content Pangaea uses the system of persistent identifiers. Four weeks after data are set to the status published, it automatically receives a DOI (Digital Object Identifier). The DOI ensures global accessibility by browsers and DOI-resolvers and can be used for exchange or referencing specific data in publications. The registration agency for data-DOI is the TIB (German National Library of Science and Technology). Data can also made citable, Pangaea being the publisher with the citation included in the library catalog of the TIB.

Data are technically formatted by the system and thus have a consistent format. This is the prerequisite for any further processing of new data compilation. Tools adjusted to the Pangaea output format allow conversion of a number of data sets to the import format of any other application.

Following OAI (Open Archive Initiative) compliant metadata are provided in xml-format for harvesting e.g. by portals or search engines. Any data set if archived in Pangaea can be found through Google.

![Diagram](image_url)

Figure 1. Pangaea plays a central role in establishing the traditional publication process for scientific data. The scientific community provides data by projects, institutes or individual scientists. Data are collected and harmonized by data curators prior to archiving in Pangaea. To ensure an Internet wide open access, the metadata to each data set are replicated in xml-format for harvesting by portals, libraries and search engines.
1.2 Data model

The great variety of parameters, methods, and interpretations used in environmental and geoscientific research are major obstacles for the integration in one common system. This was achieved by a simplified data model and a fully normalized relational database structure in combination with various middleware modules running on an application server.

The model in principle reflects the major steps in the production of georeferenced data. Through projects (PROJECT) field work, e.g. expeditions (CAMPAIGN) are carried out. During a campaign at a number of locations samples are taken or measurements are made (EVENT; the optional SITE level can be used to define a group of events). At distinct points or intervals the medium to be investigated (e.g. sediment, water or ice) is processed or subsampled (SAMPLE). Finally, from each activity or sample one to many analytical values are generated (DATA).

The main tables are supplemented by related tables of ancillary information e.g. about methods, references or parameters. Any information around the factual data (Fig. 2, green box) is called metadata, i.e. the describing data. A data set consists of a metadata header followed by one to many data series (columns), each filled with one to many data points (lines). An essential part of the model is the combination of the parameter, method, and reference tables which allows the definition of new data set types just by setting the required relations during import. Data types can be numeric or text; binary objects (in proprietary formats) are stored as files linked to a description including the URI to the object.

Each data value can be fully georeferenced in space (latitude, longitude, depth, elevation, distance) and/or time (date/time or absolute geological age); a combination of georeferences can be used as required. The value-dependent georeference allows the extraction of individual configured subsets of data and is one of the major added values, the system is providing to its users. In case a data set has millions of data points, also a data set-dependent georeferencing will be possible.

![Figure 2](image1.png)

Figure 2

The data model reflects the major steps in environmental research with projects producing georeferenced data through field campaigns and sampling events.

![Figure 3](image2.png)

Figure 3

Data can be of type numeric, text or binary objects. Each values is georeferenced in space (where ?) and time (when ?). Fields describing the method (how ?), investigator (who ?) and the parameter (what ?) complete any related metadescription.
1.3 Technical Set-up

Pangaea is based on a three tiered client/server architecture. On the client/ frontend side (1) different interfaces are offered for accessing the system. As middleware (2), an application server with components for import, retrieval and editing is operated. Finally a relational database (SQL-server/backend) is the central system with tables for the metadata and to store numeric and string values in one table each, organized through an index tree (3). Clients and SQL-server communicate via middleware running on an application server. All components are encapsulated and use standard interfaces for communication. Middleware and frontend components follow a generic model to ensure a flexible functionality and easy modifications.

Pangaea is a client/server system based on a relational database running SYBASE as the RDB management software on a multiprocessor computer. Communication via the clients is performed by open server software, organized through an application server. Communication for upload uses the ACI 4th Dimension software running on a windows server with client software for Apple OS X and Windows. A multi processor computer is used as web server for the various domains and web services. The system has an Internet connection of 1 GBit. An incremental backup runs every day, a full backup once per week. The backup is stored in two storage silos on tape, both mirroring each other and being located in two different buildings with a distance of 1 km. The system is technically operated by the computer center of the AWI.

The factual data of small to medium size data sets (< about a million data points) are stored in two tables, one for numeric values, one for string values. Large data sets or binary objects (e.g. fotos) are stored as files on harddisc or tape and are linked to its metadescription and georeference in Pangaea.
2 ACCESS
2.1 PangaVista - search and find data

The most convenient way to find data is using the search engine PangaVista. Each predefined dataset in its original granularity as defined by the PI can be found by any word matching the data set description. Keywords can be combined to create Boolean expressions using a syntax identical to those used by search engines. As a result of a query the titles of datasets are listed, each with a link for full access as a web page or download as a tab-delimited text file. The header provided with each set contains the metadata fields according to standards for describing geodata. PangaVista is not case-sensitive. Opening data sets > 20 000 data points as html may result in browser freezing.

![PangaVista search engine](image)

Figure 6. The PangaVista search engine prompts the user with a list of data sets related to a given keyword, citation with a link to the data description included.

| marine geology | Finds documents that contain both marine and/or geology. |
| marine +geology | Finds documents that contain both marine and geology. |
| "marine geology" | Placing quotation marks around any series of words turns them into a phrase and tells PangaVista that you are only interested in data sets that have the words in this specific order. |
| "marine geology" -quaternary | Finds documents that contain the phrase “marine geology” but do not contain the word quaternary. |

Table 1: Query examples for PangaVista. The syntax is similar to common search engines. Also regular expressions may be used. See URL for field names.

A map server based on the open source UMN can be used to visualize PangaVista results on a map. Data sets located at on pair of latitude/longitude are presented as red dots, data sets consisting of several geographical locations are shown as shaded blue areas. The search can be limited to a certain area by setting geographical constraints graphically or by typing in coordinates in the fields of the search engine. As geographical background information standard maps (ETOPO, GEBCO) can be used.

![Map server](image)

Figure 7: Using PangaVista’s ‘Show map’ function presents location of the data sets listed on a map. Tools allow modification of the map as required. Detailed information related to single data point can be selected by clicking onto a dot.
2.2 Accessing station metadata with PanCore

PanCore is a web-based interface, written in Java, to search through Pangaea event list and retrieve metadata on cores and samples. A map allows to set geographical constraints and is used for the visualization of hits. This client may also be used for distributing information about the availability of samples in archives. The configuration of the list allows to include metadata fields of position, elevation, time, device, etc.

Retrieval card

Starting PanCore the retrieval window is opened (Fig. 19), offering similar functionalities like PangaVista. Text and graphical retrievals are performed by filling out the form. Type words related to the core(s)/sample(s) you are looking for or open a rectangle with the mouse by pressing the left mouse button (‘Rectangle’ button activated). The coordinates of the rectangle were transferred to the coordinates window. ‘FETCH’ starts the retrieval and the data sets located within the coordinates are shown in the results window. Going back to the retrieval card the found cores are shown on the map. Zooming into the map, you can select a single core and get the detailed information by pressing the ‘Info’ button.

Configuration card

The configuration card lets you choose, which columns will be displayed in the ‘Result’ card (see Fig. 20). Highlight the appropriate metainformation item and click on the ‘Arrow’ button (or double click on the item). The order of the selected metainformation items shown in the right field of the ‘Result config.

Result card

After configuring the metainformation bring the ‘Result’ card to the top by clicking on its tab (see Fig. 21). Highlight one or more cores/samples in the list. For selecting more than one list item hold the left mouse button and move or press ‘ctrl’ and click onto the required cores/samples. Activating the ‘i’ button opens the ‘Detailed information’ window with all available metainformation. The buttons on the left side tool bar are for saving the result on your disk (txt-file), to show the complete result in a browser window, to select all items, and to get the detailed information for the current selection (see Fig. 20).
2.3 Dynamic custom retrievals with DDI

With the direct download interface (DDI) custom retrievals can be created to download specific metadata tables just with a link on a web page. The content of the list will always reflect the most recent content of the system when ever the link is used.

- Choose the table of interest.
- Define a retrieval on the 'Retrieval card'.
- Define or adjust the configuration on the 'Config card'.
- To assure that the retrieval returns the desired results test it with the 'Fetch' button.
- Save the retrieval and the configuration with the 'Save' button on the 'Retrieval' card, resp. the 'Config' card. Give the files appropriate, identical names, but keep the extensions '.retr' and '.conf'.
- You can reuse your retrieval anytime you need it with the 'Load' buttons on the 'Retrieval' card and the 'Config' card or mail it to another person (as a zip- or similar folder).
- Copy both files into a designated folder on your webserver.
- Create an url using the following syntax:

http://www.Pangaea.de/ddi?request=<url of your webfolder>/<name of your retrieval file without extension>&format=<text,textfile,html>&login=<on or off>

'request': specifies the filename without extension. 'format' (opt):
    specifies the format (text or html); textfile' requests for a filename for 'Save as' or you can add a filename between the 'ddi' and the '?' e.g. .../ddi/Filename.tab?request=... (use *.tab as extension).
'login' (opt): displays a login prompt before downloading data to enable password protection to unpublished data.

Example retrievals:

All data sets of a publication

List of samples taken from a sediment core
2.4 The ART of Data Mining

ART stands for Advanced Retrieval Tool and is a data mining tool written in Java to retrieve individually configured subsets of data from all tables of the relational data model. With ART the user is able e.g. to compare time series from different sites, parameters from different profiles, plotted on a common axis versus time or space or even allows to extract slices in time and/or space for a specific parameter. The use of ART needs some training but provides a powerful tool for the scientific work with data. A context sensitive help system is supplied.

ART should only be used by experienced users with knowledge about the Pangaea data model!

The default screen of ART to access the different tables shows a graphic of the simplified data model. Each box in the data model reflects a table of the relational system. Clicking on the box of choice will open a retrieval window, which is configured for this specific type of information. The number of cards shown depends on the tables context.

An example is given to retrieve analytical data. After clicking on Data, the user interface opens a window containing five tabs:

- (1) ‘Retrieval’ – to define the query;
- (2) ‘Config’ – configuration of the output table;
- (3) ‘List’ – listing of data in a table;
- (4) ‘Map’ – data related sites plotted on a geographical map;
- (5) ‘Plot’ – visualization of data versus space.

Figure 12. ART retrieval card.

The upper area displays the edited search expression (1). In the lower left field - subscribed ‘Criteria’ - the card shows the list of search criterions (2) that you can use in this specific context. This list will vary depending on the object that you selected in the data model. In the middle lower part a list of operators is displayed (3) and on the right are shown the values (4), that you typed in or selected from choices.

If you click on a row of the search expression (1), the components from which it is made of are highlighted resp. displayed in the lower fields. You might then change this row by selecting different items in the lower fields. Different rows of the search expression are combined by the Boolean operators ‘AND’ and ‘OR’, thus giving you the possibility to do complex retrievals. The operators are located on the left side (5). You can delete rows of the search expression by clicking on that row and subsequently clicking on the ‘Delete line’ button (6).

Values (4) are normally typed in. However, in many cases you can also use choice lists, or both, or only choices. Choice lists are useful, if you do not have an idea on how the value could be written. You can enter choices for a search criterion, if the choices button is highlighted or the add choices button is displayed. There are search criterions for which you will get immediately the complete list of choices after clicking the choices button. Other criterions, though, would result in hundreds or thousands of choice items. In this case you will get a further retrieval window, where you can first reduce the choices list by a subretrieval.
If you are using the object ‘DATA’ you can press the ‘Add’ button to add parameters to the ‘Criteria’ list (2). You will get another retrieval window, where you can do a subretrieval for the parameters that you want to retrieve data for. The subretrieval window tries to reflect the previously defined search, thus confining the parameter retrieval to what can be expected in this context. E.g., if you have defined ‘Reference is equal to publication xy’, the subretrieval window will show the same search expression and clicking on the ‘Fetch’ button brings back exactly those parameters which are relevant to this publication. Search expressions can be saved and reloaded (7).

In the ‘Options’ panel (8) you can set up the number of items to fetch (rows for metainformation, data items for DATA retrievals). On the lower part find the ‘Fetch’ button (9) to start a retrieval.

Figure 13. ART list configuration card for Event retrievals.

The configuration card presents a choice list of items to be displayed on the ‘List’ card as the final result of a search. On the left you find a choice list subscribed ‘Metainformation’ with all items that are available for display in the current context (1). On the right a default configuration is shown - named ‘Columns to display’ (2). The lists will vary depending on the object that you selected in the data model. You can add items to the default configuration either by clicking on the item and using the arrow (3) or by double clicking on an item in the left list. Items can also be removed (4). Items in the default configuration can be rearranged through the buttons ‘Up’ and ‘Down’ (5).

Each item in the default list is associated with a specific format. Text items can not be configured. For numerical items the total number of digits, the precision, and leading zeros can configured.

For date/time choose between a variety of formats (6). Use the ‘Save’ and ‘Load’ buttons (7) for custom configurations.

Figure 14. ART list configuration of a Data retrieval.

When using the retrieval for data there are two more fields to select ‘Parameters’ and ‘Geocodes’. Use the ‘Add’ button (8) to add parameters to the ‘Parameters’ list. A second retrieval window, will allow to extract the parameters of interest from the parameter dictionary. The subretrieval window may reflect the previously defined search expression on the ‘Retrieval’ card, thus confining the parameter retrieval to what can be expected in this context.

From the list of ‘Geocodes’ you can choose the type of geocoding of your data. The special geocode ‘Age calculated’ can be used for data from geological profiles, which are related to an age model. Using Agem calculate the age is calculated as part of the retrieval, based on an existing age model.

The ‘Split by events’ option (9) will return data laterally grouped by events, thus allowing a comparison of profiles between events (max 255 columns). Checking the ‘Split by versions’ option will split different versions of the same parameter. A retrieval is started by using the ‘Fetch’ button (10).
Figure 15. ART List card showing the result of a data retrieval.

List card

This card displays the result of your retrieval as a text matrix, formatted as given on the ‘List config’ card. Button have the functionality (1) to save the list, (2) to open the results in a browser window, (3) to select all rows in the list, (4) to gain more details on list items. Just click on single rows or SHIFT click to select multiple rows. Then click the ‘Details’ button and a further window shows cards with detailed information for each selected row.

Sorting of a specific column by values is enabled through clicking in the header row above that column; dragging the delimiters between columns lets you change the horizontal size of the columns.

Figure 16. ART Map card displaying the position of the retrieved data sets.

Map card

This card gives you an overview on the geographical context of the retrieved information; only available for objects having geographical coordinates.

Figure 17. ART plot card for displaying data columns in a line or scatter graph.

Plot card

If the first column of a result set contains numerical values and there is at least one further numerical column the ‘Plot’ card displays x-y-plots for each pair of numerical columns. If the values in the first column are continuously increasing or decreasing a line graph is drawn, otherwise a scatter plot.

The detail info window

(1) zoom in,
(2) geographical retrieval mode to define a geographical selection as an added constraint to the ‘Retrieval’ card with latitudes/longitudes as search criterions,
(3) detail info opens a card with detailed information for each selected point,
(4) zoom out,
(5) back to global view,
(6) map properties button opens a configuration to choose projection and center,
(7) save grafic as Postscript, (8) save as pixel grafic
Figure 18. ART detailed info window.
The detail info window shows cards with detailed information for each selected. Only one card is displayed, the others are in the background as indicated by tabs. Bring other cards in front by clicking on its tab. Click on the ‘Save’ button to download the information of all cards to your harddisk.

Global buttons are

“Fetch”: To start a retrieval. It is highlighted as soon as the search expression and the configuration is valid,

“Close”: to close the retrieval window,

“?”: to switch on context sensitive help.

Metadata example
ART retrieval example for metadata: finding gravity cores in the Mediterranean Sea:

• click on EVENT in the data model;
• click on list config tab;
• configure your table in ‘Columns to display’. You can select the searched criterions by double clicking on them and you can sort them by the up and down options. (e.g. include ‘Elevation’ and ‘Recovery’ from the metainformation menu, exclude ‘Gear type’ in columns to display);
• click on map card;
• click on rectangle in the tools bar;
• select the derived area by dragging the cursor in the global map keeping the mouse button pressed (e.g. Mediterranean Sea). The coordinates of the selected area are shown in the status line below the the map;
• click on retrieval tab (the search area is now defined);
• click on new line
• click on and
• click on ‘Gear type’ in the criteria menu and on ‘is equal to’ in the Boolean operator table; choices gives you a list of all gear types defined in the system;
• select gravity corer - wait until this field is inverted, then click ok;
• the retrieval is defined. Click on fetch; the list card gives the results, the map card shows sites;
• select i and click on one of the dots to get a more detailed information.

Save the list from the list card localy by clicking on the disk symbol in the tool bar.

Data example
Retrieving data for a specific publication (sediment surface data from Pflaumann et al. 1996)

• click on Add choices. A choices window for Reference opens. Here you enter ‘Author(s) begins with Pflau’ and press the Fetch button.
• switch to the List card in the choices window and choose (after the download is complete) ‘Pflaumann, U, Duprat, J …’ from the list and click on OK. The choices window closes and the choices list is filled with your selection (the retrieval is now defined).
• now click on the List config card. In contrast to the metainfo level there is no default configuration here. This is obvious, if you imagine that there are several thousands of parameters to choose from. You have to make a configuration:

• click the Add button above the Parameters field. A new window named Add parameter opens. In this window you can make now a subretrieval for the desired parameters. The little trick here is that the retrieval expression from the underlying DATA window is replicated, so the only thing to do is push the Fetch button. Switch to the List card, which will show all parameters which are specific for the selected publication. Choose ‘Globigerina bulloides’ and ‘Globigerina quinqueloba’ from the list (with shift click) and click OK. The Add parameter window closes and your selection is displayed in the Parameters field.
• Now you can compose your configuration. Choose for display (with double click or use arrows): ‘Event label’ as metainformation, both ‘Globigerina’ as parameters, and ‘Latitude’, ‘Longitude’ as Geocodes.
• push the Fetch button - average download time is 30 sec.

result are shown in the List card and an overview on related locations is given in the Map card. If you switch to the Map card, enable the detail info mode from the tool bar on the left side, and click on a specific point or group of points, a new window opens, displaying a detail info for each location.

Variation (last download - also the way for users of some browser platforms (see retrieval entry page) to save data on disk):

• on the Retrieval card enter ‘Fetch max. 1 data items’ and select ‘Complete results as textfile’.
• push the Fetch button again - after the download is complete, the List and Map cards will be empty, instead the results are displayed in a browser window.
2.5 The 4D-Client - curators choice

is the administrative tool for the import and editing of any information and data
to be archived in Pangaea. It enables the use of 4th Dimension operations
through a graphical user interface for the definition and update of all tables
with metainformation. Prior to the import the relations of data and metadata
are set by the curator.

Pangaea also enables the configuration of data sets by predefined retrievals
which in turn are linked to a data set description with a static URL instead of a
data set. Dynamic retrievals and the output configuration have to be
predefined either with ART or with the 4D-client.

With PangaVista, links can be defined pointing to a suite of data sets related
to e.g. a certain reference. The links can be used on external web pages or
databases to make any individual list of data available just by a mouse click.

Verwendung der klorolle und regulärer ausdrücke.
3 SOFTWARE & TOOLS

WDC-MARE serves a series of software products to be used either as standalone applications or in conjunction with the 4D-client of Pangaea. If the tools are called from the 4D-client, the software automatically starts, data are exported to the local hard disc, imported to the software and plotted in a default configuration.

3.1 PanMap

was developed for the geographical presentation of data in maps. PanMap is an easy to use geographical information system (GIS) running on Macintosh and Windows platforms. Data are organized in layers, point and vector data can be imported. Maps can be configured with different projections and individual layout. Styles of map elements can be changed, for stationary data in particular color scales can be defined according to the numerical values of related data (Fig. 3). Maps can be exported for further processing in other applications.

A collection of map resources is available on the web site, e.g. General Bathymetric Chart of the Oceans (GEBCO) \[11\] in four different resolutions, the World Vector Shoreline (WVS) \[12\] and the Global 30 Arc-Second Elevation Data Set (GTOPO30) \[13\] converted into vector data. Users can also prepare electronic data from printed maps; tools are provided for digitisation and conversion to the PanMap layer format.

3.2 PanPlot

allows the visualization of data versus time or space in standard plots or ternary diagrams (Fig. 4). Up to 255 parameter can be plotted at uniform scales. PanPlot can handle numerical, date/time, and text values, the input format is ASCII (text). Scales and graphic features can be modified and distinct parameters can be selected from the data matrix. The export of graphs is enabled in platform specific formats (EMF, PICT). Data sets exported from Pangaea can be plotted with PanPlot without conversion just by drag and drop.

![Figure 4. A ternary diagram plotted with PanPlot showing two different types of sediments.](image-url)
3.3 Ocean Data View (ODV)

[16] is a software package for the visualization of oceanographic data stored in a local database (collection). The package can be used to create and manage large sets of oceanographic data and provides tools for easy exploration and the graphical display of these data as property/property plots, color sections and color distributions on isosurfaces (Fig. 5). The collection format of ODV is optimised for dense storage and direct data access allowing the handling of a large amount of oceanographic stations on a desktop computer. ODV is distributed with the Electronic Atlas of WOCE Data (eWOCE – Schlitzer, 2000).

Figure 5. Oxygen profile through the Atlantic visualised with ODV.

3.4 Pan2Applic

3.5 PanTool

3.6 Split2Events
4.0 DATA SUBMISSION GUIDELINES
(examples are given as ID, use as http://doi.pangaea.de/10.1594/pangaea.xxxxxx)

4.0.1 general
Pangaea is a digital public library for georeferenced scientific primary data from earth and environmental research. The system does not hold model, interpolated or gridded data.
- keep consistence - no abbreviations - avoid redundancies -
- use the decimal point - use the English language - check for misspelling.

4.0.2 metadata
- Keep campaign and event labels unique, short and without blanks (use _); use proper labels (e.g. as given in the cruise report); provide lists of new events using the form at http://www.Pangaea.de/curator/files/InpEvent.xls; event labels in Pangaea can be checked using the PanCore client at http://www.Pangaea.de/PanCore.
- Valid formats for positions (latitude/longitude) are: 65°45’32”S or 65°45.7’S or -65.1234 (South and West are negative when using decimal degree).
- Elevation is negative below sea level and positive above sea level.
- For the data set citation, provide author(s) and a title (equivalent to a title of publications); when using PangaVista a user should easily understand the content of a data set (who has analyzed when, where, what and how?); the title should contain e.g. a short description of the data, the geographical area and the sample type, 50184.
- Provide related reference(s) complete and with DOI (if applicable).
- PI (principle investigator, with email) should be given for each parameter; the method may be added for all parameters where applicable.
- Add any comments or descriptions needed to understand a data set; comments can be given per event, per value, per data series, and per data set; for data sets provide text unformatted (~ half page) or in PDF/A-format (> half page).
- New parameters are defined by the data librarian (hgrobe@pangaea.de); for already available parameters with ID see http://www.Pangaea.de/parameter. Clearly separate method from parameter when defining a new parameter.

4.0.3 data
- The first column of a data file always is headed by the expression 'Event label'; columns with geocode(s) and data to follow.
- For data, the header line must contain the full parameter name(s) with unit in squared brackets or its abbreviation or its Pangaea-ID. Examples:
  - Temperature, water [deg C] or Temp [deg C] or 717
  - Homo sapiens or H. sapiens or 16567
  - Calcium carbonate [%] or CaCO3 [%] or 70
- Columns with numerical parameters must contain numbers only (exception: quality flags may head numbers, see 4.3).
- Precision, i.e. number of digits after the decimal point, must express accuracy of method and results.
- Cells in columns with text parameter can be filled with a max. length of 255 characters.
- Lines and columns which are empty or should not be imported, must be removed.
- An empty cell in a data matrix means 'not analyzed'. If the result of the analysis is zero, the cell must be filled with a '0' and not left blank (78037). In tables using text parameters, a dash (-) can be used instead (63404).
- Provide tables for import in tab-delimited ASCII text files or as excel file without formatting and macros (data files are processed and imported to a relational database, thus any formatting will be discarded). Metadata related to data may be submitted as unformatted text. Submit corresponding files in one zip-archive.
- It is useful to combine data files of many events with identical parameters in one file with parameters in columns and events in lines; if a collection consists of many single files, it must be ensured, that all files have the same format.
- The import (and export) format for date/time is ISO8601: YYYY-MM-DDThh:mm:ss.
- The import format for latitude and longitude in data files is decimal degree.
- If a parameter is imported more than once in one data set, the difference must be visible through the method, PI and/or the data series comment (78746).
- Technical data quality and consistence should be checked prior to upload.
- Scientific data quality is in the responsibility of the author(s) and PI.
- ALWAYS check data after import carefully using http://www.pangaea.de/PangaVista.
- Send data set DOI or links to PI/author for proof read and corrections.
- Password protection for a limited time (moratorium) is in the decision of the project or PI and is defined on request.
4.1 Metadata and data preparation for upload

Data are imported and edited with the 4D-client according to the hierarchical structure of the Pangaea. The data model consists of four main tables:

Project > Campaign > Event > Data.

Prior to the import of primary (factual) data, always the metadata need to be defined. Relations between data and metadata are set by the import routine.

Project

Ensure the project is defined, otherwise request for a new definition via email to the data librarian consisting of the full title, acronym, coordinator and URL of the projects homepage.

Campaign (synonym: cruise/expedition/leg)

Metainformation about expeditions is stored in the campaign table. This level is optional; it enables the user to find events and/or data related to a certain campaigns. Define a new campaign by typing on an empty line in the list of campaigns in the project window or use the form at http://www.pangaea.de/curator/files/InpCampaign.xls for the import of a list of campaigns.

Event (synonym: station/site/observation/core)

The fields event label, latitude and longitude are mandatory. The event label must be unique within Pangaea. This can be ensured by combining the station number with the campaign label (e.g. Taymyr97_3). Several events can be grouped on the site level. Single events can be defined starting from the related campaign. A list of events is imported using the form http://www.pangaea.de/curator/files/InpEvent.xls.

Metadata related to the factual data

- Define authors and PI of the data to be imported in the staff table.
- Define or import related references in the reference table.
- Define methods related to the parameters in the method table.
- Ensure, that documentation files to be linked to the data set description in the further details field are stored on the web server and the link is known.

Data (of analysis/observations/measurements)

In principle the granularity is defined as one to many data sets per event. A set should consist of e.g. all parameters belonging to one analysis or method.

Examples of typical data set ‘granules’ are:

- Physical oceanography from a CTD (71590)
- Grain size distribution of a sediment core (91227)
- Photos from one sea-bed rov station (227670)
- Mineralogy of surface sediments (57249)
- Isotopes measured in a biological object (54254)
- Oxygen lander profile through the water/sediment interface (198921)
- Time series with oceanographic measurements over one year (55685)
- Seismic profile (206529)

An import data set in principle consists of parameters in columns and samples in lines with the

1. column containing 'Event label' in header, followed by the label as defined in the event table.

2. (- 3.) column containing the third spatiotemporal dimension (geocode):

- DEPTH, sediment [m] (1)
- DEPTH, water [m] (1619)
- DEPTH, ice [m] (5059)
- DISTANCE [cm] (2920) in an object, e.g. coral, tree, shell, nodule
- ELEVATION [m] (8128) for samples taken on the earth surface, mostly above sea level
- ALTITUDE [m] (4607) for measurements or samples in the air
- ORDINAL NUMBER (6262) can be used, if no third dimension is given and samples are just listed; may be used to force a data set appearing in a given order.
- A data set can only have ONE vertical spatial geocode but may ALSO have a time geocode:
  - DATE/TIME (1599) for time series (in ISO-format: YYYY-MM-DDThh:mm:ss)
  - AGE [kyr] (2205) for the calculated age of each sampling/measurement point; for ages, resulting directly from dating methods (AMS, Lead210, OSL etc) use parameter Age, dated (5506). Age models resulting from the interpretation of datings are imported with ID=6 or 6666 (Age model) see also 4.1.1.

The following columns contain the data, each with its parameter name in the header line.

Open the data file to be imported with the 4D-client by using Menue/import/analytical data. Ensure that all parameters are recognized by the system. Define relations between data set and metadata. After import check correct upload, data-metadata relations and formats via Pangavista.
4.1.1 The final AGE solution

Each sample, used for age determinations in sediments, has one mean depth or, even better two depth, defining the upper and lower boundary of the sampled horizon. At a minimum an import table for datings should include (ID in brackets)
- Depth (1), mean
- Depth (3+4), min + max (optional)
- Age, dated in kyr (1506)

It is important to define the correct relation between Age and the Method of dating which can be found by give dynamik link list: method begins with “Age,...

if 14C data are archived, always
- specify the calibration used for converting 14C-years to calendar years in the method (e.g. INTCAL 98, Suigetsu)
- specify the reservoir correction used for marine 14C dates

For each age a confidence level may be added, given through the parameters
- Age, min (UCL, upper confidence limit, 6170)
- Age, max (LCL, lower confidence limit, 6169)

if the limits are different. Otherwise use Age, dated standard deviation (6165). All age parameters have the unit kiloyears (kyr)!

Age, dated material (2836) is a text parameter to add a description of the material which was used for the age determinations. Add parameter Age, comment (643) for any comments needed.

If data are provided versus year A.D. or counted years (e.g. varves, growth bands, tree rings), the year is kept in the parameter Year (2936, unit yr A.D.) and the geocode AGE converted to kyr related to 1950=0 is added.

Following the scientific decision, which age determinations are usefull, an age model is set up consisting of a number of stratigraphic fixpoints. Use parameter Age model (6) or, if more age models exist, an optional age model parameter is available (6666). Age, comment gives information why a certain point/age was choosen. To precisely define the upper and lower limit of an age, the system provides
- the numeric parameter Age, minimum/young (6167)
- and the text parameter Age from (oldest) (45397)
- Age to (youngest) (45398)

Also with age models it is important to define the correct relation between Age model and the method used; see give dynamik link list: method begins with “Age model...

A Hiatus is defined by using its depth twice with upper and lower time limit of the hiatus. Data from geological profiles should consequently be accompanied by its age model.

From the age model, the ages of the distinct samples are calculated (by linear interpolation or other).

Never use the parameter Age, calculated for import ! Age, calc is used for the export of calculated ages only by using the values in the age model.

The geocode AGE should be included with all data related to geological time. It is calculated for the age model. AGE is the common time axis of the system, allowing to combine any type of geological data, e.g. corals, ice, sun spots and reconstructed sea surface temperatures.

A complete list of parameters with ID used for ages is given by give dynamik link list of age parameter

<table>
<thead>
<tr>
<th>Geob1003-1</th>
<th>Event label</th>
<th>m</th>
<th>6</th>
<th>1234 estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GeoB1003-1</td>
<td>0.0</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.34</td>
<td>12.5</td>
<td>termination 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.25</td>
<td>18.2</td>
<td>upper limit of hiatus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.25</td>
<td>35.0</td>
<td>lower limit of hiatus</td>
</tr>
</tbody>
</table>
4.2 Work flow during data publication

data should be made available latest when a publication is accepted;

- data are provided by the author/PI, who is responsible for
  - (1) validation and scientific quality control,
  - (2) preparation of import format,
  - (3) completeness of metadata,
  - (4) submitting the data to the data curator
- the data curator is responsible for
  - (1) checks of metadata,
  - (2) technical quality control,
  - (3) import to Pangaea
  - (4) submit DOIs to PI
- proof read of data sets by PI, send corrections to curator
- corrections by curator
- publication (after a moratorium on request)

Each data sets is published with a full citation consisting of
- the list of authors,
- the title,
- the source,
- the publisher (i.e. Pangaea), and
- a technical DOI.

On request a publication DOI can be defined for single sets or groups of sets, including a reference in the library catalog of the TIB (http://tws.gbv.de).

4.3 Quality flags

Each numeric value of an import file can be quality flagged with symbols as listed below. Insert the symbol at the beginning of a value without a blank in between. Values without any symbol are qualified as being valid.

<table>
<thead>
<tr>
<th>Quality flag</th>
<th>Symbol</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>valid</td>
<td>(default)</td>
<td>2.3</td>
</tr>
<tr>
<td>questionable</td>
<td>?</td>
<td>?9.1234</td>
</tr>
<tr>
<td>not valid</td>
<td>/</td>
<td>/21.21</td>
</tr>
<tr>
<td>unknown</td>
<td>*</td>
<td>*198</td>
</tr>
<tr>
<td>custom</td>
<td>#</td>
<td>#77.7</td>
</tr>
<tr>
<td>larger than</td>
<td>&gt;</td>
<td>&gt;14000</td>
</tr>
<tr>
<td>less than</td>
<td>&lt;</td>
<td>&lt;0.05</td>
</tr>
</tbody>
</table>

When using Pan2Applic quality flags are converted to those used by the application, e.g. when using ODV. (See also documentation of project specific quality flags at http://www.pangaea.de/Projects/WOCE/woce_quality_flags.pdf.)
4.4 Login & access rights

Pangaea has three user levels: administrator, curator and user.

In PangaVista any metainformation and all published data can be accessed without a login by any user.

Unpublished data can be accessed by the PI and administrator, unless defined for other users or groups with the 4D-client.

ART can be used for data retrieval and download by curators and administrators only.

An account is required to use the 4D-client.

Administrator is allowed to change any information and to define projects and parameter.

Curator can change meta-information which was imported under their account.

Access of individual users or groups to unpublished data sets can be defined with the access right button on the data set level.

To define access to unpublished data:

- select the relevant data sets,
- click on the access-right button,
- drag and drop a user or a group of users on the data set(s),
- the new definition of access rights will be listed in the upper left Rights' window,
- press 'Save'.

Data sets never consist of both, published and unpublished data.

User name and password are identical in 4D- and in the web-clients.

New users are defined on request by rsieger@Pangaea.de.

Passwords can be changed through the password requester.

Within a project unpublished data should be shared through a common project account.

4.5 Templates and examples
**Sediment profile (cores or outcrops)**

use for any kind of vertical sampling or measuring profile through a sediment section, e.g. from cores or outcrops and for any group of parameters which can be determined in the sediment, e.g. microfossils, granulometry, chemistry, physical properties

**Metadata of location**
(site/station/event/sampling point)
need to be defined first; provide as list, e.g. station list of cruise; use form InpEvent.xls

mandatory: event label, latitude, longitude
recommended: elevation below(-) or above (+) sea level, device, campaign, date, time

optional: optional label, area, comment

**Metadata of data**
mandatory: author(s), title of data set, PI for each parameter
recommended: method, related reference(s), project(s)

optional: any comments to data series and/or data set

**Header line of import data matrix**
1. column: Event label
2. column: unit used for depth in sediment: ‘m’ or ‘cm’
3. to n. column: parameter [unit] or its ID in Pangaea

For ID of parameters used in sediment investigations see http://www.pangaea.de/curator/parameter.html

If several profiles with the same parameters are to be imported, all profiles can be provided in one matrix. The event label may appear in the first line only or in each line.

---

**Matrix ready for import:**

<table>
<thead>
<tr>
<th>Event label</th>
<th>m</th>
<th>67</th>
<th>Sand [%]</th>
<th>d13C [per mil]</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS24/013-5</td>
<td>0</td>
<td>1.209</td>
<td>23.5</td>
<td>-22.13</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>1.117</td>
<td>12.5</td>
<td>-21.34</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>1.032</td>
<td>67.0</td>
<td>-21.92</td>
</tr>
<tr>
<td></td>
<td>2.6</td>
<td>0.345</td>
<td>44.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.85</td>
<td>0.765</td>
<td>56.0</td>
<td>-21.1</td>
</tr>
<tr>
<td>CD123-456</td>
<td>0.55</td>
<td>1.1</td>
<td>9.1</td>
<td>-22.34</td>
</tr>
<tr>
<td>CD123-456</td>
<td>10</td>
<td>2.34</td>
<td>6.0</td>
<td>-21.07</td>
</tr>
<tr>
<td>CD123-456</td>
<td>20.1</td>
<td>3.21</td>
<td>5.2</td>
<td>-23.0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example data set for export** doi:10.1594/Pangaea.51160
Pollen profile

Metadata of location
(Event label ⇒ site/station/event/sampling point)
The event label needs to be defined first; if several, provide as list, e.g. station list of expedition; use form InpEvent.xls
mandatory: event label, latitude, longitude
recommended: elevation below(-) or above (+) sea level, device,
campaign, date, time
optional: optional label, area, comment

Metadata of data
mandatory: author(s), data set title, PI for each parameter
recommended: method, related reference(s), project(s)
optional: any comments to data series and/or data set as free text (a few lines) or as pdf-file (if more)

Header line of import data matrix
1. column: Event label
2. column: unit of depth in profile used > ‘m’ or ‘cm’
3. to n. column: parameter [unit] or parameter abbreviation [unit] or parameter ID as defined in Pangaea

Matrix ready for import:

<table>
<thead>
<tr>
<th>Event label</th>
<th>cm</th>
<th>Typha latifolia [#]</th>
<th>Lyc.a [#]</th>
<th>1234</th>
<th>56789</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalkriese</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>2</td>
<td>14</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2.6</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>7.85</td>
<td>7</td>
<td>56</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>10.3</td>
<td>2</td>
<td>33</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>12.7</td>
<td>12</td>
<td>123</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>20.1</td>
<td>8</td>
<td>345</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Example how a pollen profile appears if archived in Pangaea:
doi:10.1594/Pangaea.142610

For ID of pollen parameter see http://www.Pangaea.de/curator/parameter.html
If several profiles with the same parameters are to be imported, all profiles may be provided in one matrix. Repeating event label in each line is optional.
Hydrographic profile

use for any kind of vertical sampling or measuring profile through the water column, e.g. CTD/Rosette, Multinet ...)

Metadata of location
(Event label ⇒ site/station/event/sampling point)
need to be defined first; if several, provide as list, e.g. station list of expedition.
mandatory: event label, latitude, longitude
recommended: water depth (negative, e.g. 10 m depth=-10),
device, campaign, date, time
optional: optional label, area, comment

Metadata of data
mandatory: author(s), data set title, principle investigator (PI) for each parameter
recommended: method, related reference(s), project(s)
optional: any comments to data series and/or data set as free text (a few lines) or as pdf-file (if more)

Header line of import data matrix
1. column: Event label
2. column: ID for depth in water: 1619
further columns: parameter [unit] or
parameter abbreviation [unit] or
parameter ID as defined in Pangaea

Matrix ready for upload:

<table>
<thead>
<tr>
<th>Event label</th>
<th>Press [dbar]</th>
<th>Cond [mS/cm]</th>
<th>Salinity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS58/123-4</td>
<td>1.98</td>
<td>-1.7235</td>
<td>26.1254</td>
</tr>
<tr>
<td></td>
<td>3.96</td>
<td>-1.7067</td>
<td>26.386</td>
</tr>
<tr>
<td></td>
<td>5.95</td>
<td>1.6824</td>
<td>26.6026</td>
</tr>
<tr>
<td></td>
<td>7.93</td>
<td>1.5833</td>
<td>26.9876</td>
</tr>
<tr>
<td></td>
<td>9.91</td>
<td>1.5987</td>
<td>26.9876</td>
</tr>
<tr>
<td></td>
<td>11.89</td>
<td>1.5345</td>
<td>26.1234</td>
</tr>
<tr>
<td></td>
<td>13.88</td>
<td>1.5</td>
<td>26</td>
</tr>
</tbody>
</table>
| ...         | ...          | ...          | ...      | ...
| PS58/456-7  | 2            | 0.5678       | 24.543   | 35.654   |
| ...         | ...          | ...          | ...      | ...

Example of a hydrographic profile data set:
doi:10.1594/Pangaea.226732

For ID of hydrographic parameter see [http://www.Pangaea.de/curator/parameter.html](http://www.Pangaea.de/curator/parameter.html)
If several profiles with the same parameters are to be imported, all profiles may be provided in one matrix. Repeating event label in each line is optional. If only pressure is given, the water depth must be calculated by using the factor 1.025 (water depth = pressure x 1.025).
Time series

use for any kind of a time series of measurements

Metadata of location
(Event label ⇒ site/station/event/sampling point)
need to be defined first; if several, provide as list, e.g. station list of expedition.
use form http://www.Pangaea.de/curator/files/InpEvent.xls

mandatory: event label, latitude, longitude
recommended: water depth, device, campaign,
date, time (start of time series),
date2, time2 (end of time series)
optional: optional label, area, comment

Metadata of data
mandatory: author(s), data set title, principle investigator (PI) for each parameter
recommended: method, related reference(s), project(s)
optional: any comments to data series and/or data set as
free text (a few lines) or as pdf-file (if more)

Header line of import data matrix
1. column: Event label
2. column: Date/Time: 1599 (given in ISO format)
3. column: ID for 3. Geocode (if applicable)
further columns: parameter [unit] or
parameter abbreviation [unit] or
parameter ID as defined in Pangaea

Matrix ready for upload:

<table>
<thead>
<tr>
<th>Event label</th>
<th>1599</th>
<th>1619</th>
<th>717</th>
<th>Salinity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SurfClam</td>
<td>2004-04-12T12:38</td>
<td>5</td>
<td>-1.7235</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2004-04-12T13:38</td>
<td>5</td>
<td>-1.7067</td>
<td>32.99</td>
</tr>
<tr>
<td></td>
<td>2004-04-12T14:38</td>
<td>5</td>
<td>1.6824</td>
<td>33.245</td>
</tr>
<tr>
<td></td>
<td>2004-04-12T15:38</td>
<td>5</td>
<td>1.5833</td>
<td>33.745</td>
</tr>
<tr>
<td></td>
<td>2004-04-12T16:38</td>
<td>5</td>
<td>1.5987</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2004-04-12T17:38</td>
<td>5</td>
<td>1.5345</td>
<td>34.12</td>
</tr>
<tr>
<td></td>
<td>2004-04-12T18:38</td>
<td>5</td>
<td>1.5</td>
<td>34.543</td>
</tr>
<tr>
<td></td>
<td>2004-04-12T19:38</td>
<td>5</td>
<td></td>
<td>34.6</td>
</tr>
<tr>
<td></td>
<td>2004-04-12T20:38</td>
<td>5</td>
<td>0.5678</td>
<td>35.654</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Example of a Time series data set:
doi:10.1594/Pangaea.308901
Tree rings

Metadata of location
(Event label ⇒ site/sampling point/lab or series ID)

use form InpEvent.xls
mandatory: event label, latitude, longitude
recommended: elevation above sea level [m], sample type,
date, time (of sampling)
optional: optional label= tree code, geographical area,
campaign, comment (Waldkante, Splint, Mark,
Ringzahl, Start- and Endjahr)

Metadata of data
mandatory: author(s), data set title, PI for each parameter
recommended: method, related reference(s) if published,
project(s)
optional: any comments to data series and/or data set as
free text (a few lines) or as pdf-file (if more)

Header line of import data matrix
1. column: event label
2. column: Age [kyr]
3. column: Year AD
4. column: Ring width [1/100 mm]
5. column: number of trees used (optional)

If several profiles with the same parameters are to be imported, all profiles may be provided in
one matrix. Repeating event label in each line is optional.

Matrix ready for upload:

<table>
<thead>
<tr>
<th>Event label</th>
<th>Age [kyr]</th>
<th>Year AD</th>
<th>Ring Width [1/100 mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>fag1234</td>
<td>1920</td>
<td>2</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>1921</td>
<td>4</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>1922</td>
<td>6</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>7.93</td>
<td>8</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>9.91</td>
<td>10</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>11.89</td>
<td>12</td>
<td>71</td>
</tr>
<tr>
<td>fag9876</td>
<td>13.88</td>
<td>14</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>15.86</td>
<td>16</td>
<td>95</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Example how a hydrographic profile appears for download if archived in Pangaea: doi:10.1594/Pangaea.60032
**Horizontal profile (generic)**

(large volume data from e.g. geophysics or bathymetry)

**Metadata of location**

(for import of a list of profiles use form InpEvent.xls)

**mandatory:** event label (profile identifier),
latitude, longitude (start & end)

**recommended:** profile type, date, time (start & end)

**optional:** elevation, geographical area, campaign, comment

**Metadata of data**

**mandatory:** author(s), data set title, principle investigator (PI),

**recommended:** method, related reference(s) if published, project(s)

**optional:** any comments and descriptions as free text

(< 1/2 page) or as pdf-file (if more)

Data are archived as files in a storage system. The data set to be imported consists of a description of the profile in date/time and lat/long with columns containing the link to the related data sets.

The granularity of a data set needs to be defined in a scientifically useful form and in a size which can still be downloaded through the Internet (< 100 MB).

**Header line of import data matrix**

1. column: Event label
2. column: Date/time (1599) in ISO-format
3. column: Latitude (1600) in decimal degree
4. column: Longitude (1601) in decimal degree
5.-n. column: technical and service parameter
e.g. speed, direction, elevation

n.+1 column: URL (15651)
**Refraction seismic profile with OBS**

**Metadata of location**
use form InpEvent.xls for list of all seismic stations on one profile
mandatory: event label (of seismometer), latitude, longitude
recommended: elevation, device type, date, time
optional: optional label, geographical area, campaign, comment

**Metadata of data**
mandatory: author(s), data set title, principle investigator (PI)
recommended: method, list of reference(s) if published (may include the report of the campaign), project(s)
optional: any comments and descriptions as free text
(< 1/2 page) or as pdf-file (if more)

**Header line of import data matrix**
1. column: Event label
2. column: URL (15651)
3.-n. column: technical and service parameter as required (optional)

The links to the OBS data files from one profile are imported together in one data set. OBS data files are stored in SGY-format in a storage system.

**Matrix ready for upload:**

<table>
<thead>
<tr>
<th>Event label</th>
<th>URL</th>
<th>Ring Width [1/100 mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS56/123-4</td>
<td><a href="http://hs.Pangaea.de/gwp/">http://hs.Pangaea.de/gwp/</a>...</td>
<td>43</td>
</tr>
<tr>
<td>fag9876</td>
<td>13.88</td>
<td>123</td>
</tr>
<tr>
<td>...</td>
<td>15.86</td>
<td>95</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Example of a link to a single OBS data set: doi:10.1594/Pangaea. 269574
Reflection seismic profile

Metadata of location
use form InpEvent.xls for list of all seismic stations on one profile
mandatory: event label (of seismometer), latitude, longitude
recommended: elevation, device type, date, time
optional: optional label, geographical area, campaign, comment

Metadata of data
mandatory: author(s), data set title, principle investigator (PI)
recommended: method, list of reference(s) if published (may include the report of the campaign), project(s)
optional: any comments and descriptions as free text (< 1/2 page) or as pdf-file (if more)

Header line of import data matrix
1. column: Event label
2. column: URL (15651)
3.-n. column: technical and service parameter as required (optional)

The links to the OBS data files from one profile are imported together in one data set. OBS data files are stored in SGY-format in a storage system.

Matrix ready for upload:

<table>
<thead>
<tr>
<th>Event label</th>
<th>URL</th>
<th>Ring Width [1/100 mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS56/123-4</td>
<td><a href="http://hs.Pangaea.de/gwp/">http://hs.Pangaea.de/gwp/</a>...</td>
<td>43</td>
</tr>
<tr>
<td>fag9876</td>
<td>13.88</td>
<td>123</td>
</tr>
<tr>
<td>...</td>
<td>15.86</td>
<td>95</td>
</tr>
</tbody>
</table>

Example of a link to a single OBS data set: doi:10.1594/Pangaea.269574
Gravimetric profile

Metadata of location

use form InpEvent.xls for list of all seismic stations on one profile

mandatory: event label (of profile),
latitude, longitude, elevation, date, time (start),
latitude, longitude, elevation, date, time (end),
sonar type

recommended: elevation, device type (=Gravimeter), date, time

optional: optional label, geographical area, campaign,
comment

Metadata of data

mandatory: author(s), data set title, principle investigator (PI)

recommended: method, list of reference(s) if published (may
include the report of the campaign), project(s)

optional: any comments and descriptions as free text
(< 1/2 page) or as pdf-file (if more)

Header line of import data matrix

1. column: Event label
2. column: URL (15651)
3.-n. column: technical and service parameter as required
(optional)

The links to the OBS data files from one profile are imported together in one
data set. OBS data files are stored in SGY-format in a storage system.

Matrix ready for upload:

<table>
<thead>
<tr>
<th>Event label</th>
<th>URL</th>
<th>Ring Width [1/100 mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Track.SO175</td>
<td><a href="http://hs.Pangaea.de/gwp/">http://hs.Pangaea.de/gwp/</a>...</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>13.88</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>15.86</td>
<td>95</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

parameter: gravity (mgal), ascii, 1 file/leg, resolution 1 sec?

Example of a link to a single OBS data set: doi:10.1594/Pangaea.269574
**Magnetic profile**

**Metadata of location**

use form InpEvent.xls for list of all seismic stations on one profile

mandatory: event label (of seismometer), latitude, longitude

recommended: elevation, device type, date, time

optional: optional label, geographical area, campaign, comment

**Metadata of data**

mandatory: author(s), data set title, principle investigator (PI)

recommended: method, list of reference(s) if published (may include the report of the campaign), project(s)

optional: any comments and descriptions as free text (< 1/2 page) or as pdf-file (if more)

**Header line of import data matrix**

1. column: Event label
2. column: URL (15651)
3.-n. column: technical and service parameter as required (optional)

The links to the OBS data files from one profile are imported together in one data set. OBS data files are stored in SGY-format in a storage system.

---

**Matrix ready for upload:**

<table>
<thead>
<tr>
<th>Event label</th>
<th>URL</th>
<th>Ring Width [1/100 mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS56/123-4</td>
<td><a href="http://hs.Pangaea.de/gwp/">http://hs.Pangaea.de/gwp/</a>...</td>
<td>43</td>
</tr>
<tr>
<td>fag9876</td>
<td>13.88</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>15.86</td>
<td>95</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

parameter: x,y,z,F, ascii, 1 file/leg, resolution 1 sec ?

Example of a link to a single OBS data set: doi:10.1594/Pangaea.269574
**Bathymetric profile (swath sonar)**

**Metadata of profile**

use form InpEvent.xls for list of all profiles

mandatory: event label (of profile),
latitude, longitude, elevation, date, time (start),
latitude, longitude, elevation, date, time (end),
sonar type

optional: optional label, geographical area, campaign,
comment

**Metadata of data**

mandatory: author(s), data set title, principle investigator (PI)

recommended: sonar type with specification, list of reference(s) if published (may include the cruise report), project(s)

optional: any comments and descriptions as free text (< 1/2 page) or as pdf-file (if more)

**Header line of import data matrix**

1. column: Event label
2. column: Date/time (1599) in ISO-format (10 min step)
3. column: Latitude (1600) in decimal degree
4. column: Longitude (1601) in decimal degree
5.-n. column: technical and service parameter
e.g. speed, direction, center beam depth
n.+1 column: URL (15651) to binary objects and xml-file

**Matrix ready for upload:**

<table>
<thead>
<tr>
<th>Event label</th>
<th>URL</th>
<th>Ring Width [1/100 mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS56/123-4</td>
<td><a href="http://hs.Pangaea.de/gwp/">http://hs.Pangaea.de/gwp/</a>...</td>
<td>43</td>
</tr>
<tr>
<td>fag9876</td>
<td>13.88</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>15.86</td>
<td>95</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Granularity of data is max. 8 hours per data set. Time resolution will be about one second. Data are stored as preprocessed raw data in SURF-Format (STN Atlas). Other format options are still under discussion (Triton-Eics XTF, Caris HDCS, Generic Sensor Format (GSF),...). Metadata files are added in XML-format for each data set. Final processed data are archived in phi/lambda/depth in ASCII.

Example of a swath sonar profile: doi:10.1594/Pangaea. xxxxx
ADCP profile

Metadata of profile
use form InpEvent.xls for list of all profiles

mandatory:
- event label (of profile),
- latitude, longitude, elevation, date, time (start),
- latitude, longitude, elevation, date, time (end),
- sonar type

optional:
- optional label, geographical area, campaign,
- comment

Metadata of data

mandatory:
- author(s), data set title, principle investigator (PI)

recommended:
- sonar type, list of reference(s) if published (may include the cruise report), project(s)

optional:
- any comments and descriptions as free text (< 1/2 page) or as pdf-file (if more)

Header line of import data matrix

1. column: Event label
2. column: Date/time (1599) in ISO-format (10 min step)
3. column: Latitude (1600) in decimal degree
4. column: Longitude (1601) in decimal degree
5.-n. column: technical and service parameter
e.g. speed, direction, center beam depth
n.+1 column: URL (15651)

---

Matrix ready for upload:

<table>
<thead>
<tr>
<th>Event label</th>
<th>URL</th>
<th>Ring Width [1/100 mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS56/123-4</td>
<td><a href="http://hs.Pangaea.de/gwp/">http://hs.Pangaea.de/gwp/</a>...</td>
<td>43</td>
</tr>
<tr>
<td>fag9876</td>
<td>13.88</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>15.86</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

parameter ? granularity ?

Example of a swath sonar profile: doi:10.1594/Pangaea.xxxxx