BEAM

Blacknest Data System (BDS)
User Manual – 1.1.0

Project	Blacknest
Date	23/07/09
Reference	blacknest/bdsUserManual
Author	Dr Terry Barnaby

Table of Contents

BEAM

1. References

- The BEAM Blacknest support website at: https://portal.beam.ltd.uk/support/blacknest. This provides detailed information on the BDS system and the current AutoDRM, information on alternative AutoDRM implementations and information on data formats.

2. Introduction

This document provides an overview of the Blacknest Data System (BDS). The purpose of the system is to provide storage and access to seismic data and associated information. Development of the BDS system is an on-going process adding features and working around issues in the seismic data and meta data as they arise.

Seismic data is obtained from seismic stations and arrays of stations situated at various places in the world. Each seismic station consists of one or more seismometers detecting earth movement in one or more directions. The low frequency analogue data from each seismometer is digitally sampled, at a low sample rate, and stored as a channel in a data file. There is also an archive of analogue data on tapes which will be digitised for inclusion into the system.

The BDS system provides a core platform that is modular in design and integrates the seismic sensor data with its meta data.. The BDS core defines API's and data formats that can be used and an overall system structure. The basic core system can be extended with additional modules as required.

2.1. Original Data

The original seismic sensor data is stored in files in a number of different data formats. In general each data format contains a header, providing information on the contained data, such as the time, the array/station name, the channel and the sample rate. There are then one or more channels of time domain samples from the array's or station's seismometers. The definition of which instrument the channels data is from is either in the header, in a separate Instrument Response Database, in various paper documents and sometimes unknown. Blacknest has a large robotic tape data storage archive containing a large set of seismic sensor data from local as well as remote seismic arrays and a preliminary MySQL database containing the Instrument Response information.

2.2. BDS Features

The systems core features are as follows:

1. To archive seismic sensor data on data storage systems.

2. To archive seismic meta data such as instrument settings and calibration information.

3. To index the seismic data and meta information stored in the system.

4. To provide access to the data in the data storage system both locally and remotely.

5. To provide data format converters to allow the data to be retrieved in the format required.

6. To provide various data front-ends including an Autodrm, command line and Web based interfaces.

BEAM

7. Can use multiple data storage systems such as tape archive and disk archive systems.

8. To provide data input systems: command line, GUI and automatic Scream Satellite input.

9. Data access security by userid/password.

10. To provide access to data from real-time incoming data feeds. (Not in version 1)

11. Bandwidth management for remote links. (Not in version 1)

12. Uses a RAID disk array as the primary data storage archive.

13. Provides a 'C++' API for easy and direct access to the data.

Future system features could include:

1. Provide additional program API's to allow easy and direct access to the data. This could be implemented for: Python, Octave, MatLab, (SAC, GeoTool), etc.

2. Provide in-built algorithms for data manipulation. This would allow, for example, sample rate changing etc.

3. Provide robot programs to look for data features etc.

4. Data backup ability to remote archive. This could just be the data base and meta-information but could also include the main data.

5. Ability to feed data to other sites

6. Event system for errors such as satellite data feed errors.

2.3. Notes

- The data access front-ends are separated from the data gathering and data conversion systems. This allows other front ends to be easily added and modularises the system. Examples of different front ends are: email, web and direct access from local data processing programs.

- The data converters are implemented as separate modules. This allows them to be used for other purposes and allows for better testing of their functionality. It also allows other data converters to be added to the system easily. The data converters convert between their own format and a special internal format. This internal format provides a generic API that can handle the data present in all data formats. The BDS systems uses its own internal data format for storing data on disk. It is not intended to be an external published format so that it can be easily changed to suit internal requirements.

- The system uses a MYSQL Database for storing the user, indexing and instrument changes and calibration information.

- The method used to retrieve the data is implemented in a modular way. This modularity allows other data access modules to be added as required in addition to the standard disk archive access method.

- The system has the concept of security groups. This allows certain users to access data that is not accessible to all.

- Full logging and statistical information is available.

- A password protected, administration interface is provided for ease of administration and the viewing of major system statistics.

BEAM

3. Design Overview

This section describes the overall BDS design. The design is modular and allows extension as well as use of its components by external systems. The fundamental split is between the input/output modules and the main Data Processing Engine (BDSE).
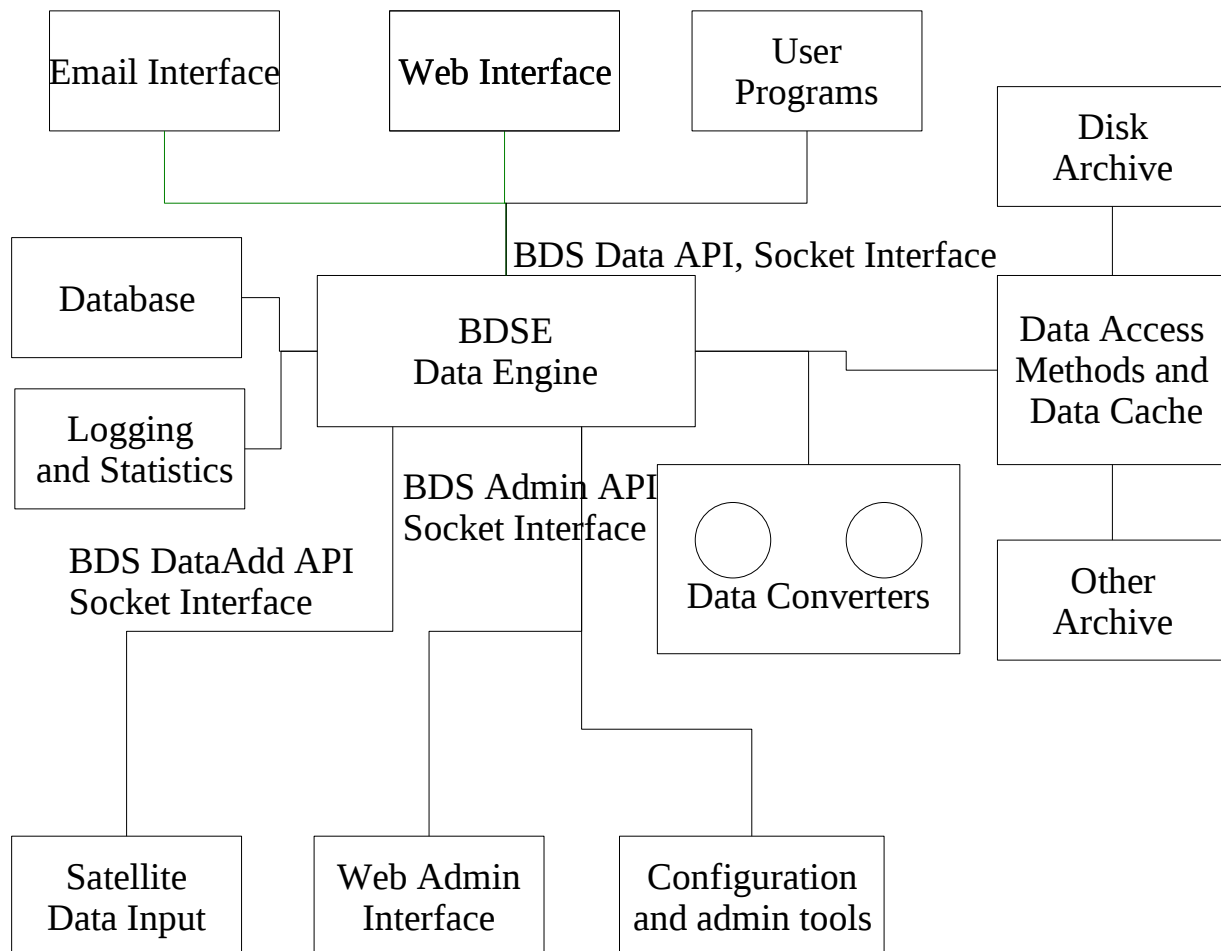
BDS Overall Design

The BDS modules communicate with the main "Data Engine" using the BOAP Object Orientated RPC protocol over an network socket interface.

The "Data Engine" (BDSE) is the heart of the system. It is implemented in the "bdsServer" program. It queues requests for data, prioritises them, gathers the data and converts it into the appropriate format using the converters available. All accesses to read or write data go through the BDSE. It is responsible for keeping the data protected and valid.

It provides three separate API's. These have independent protected access:

- BDS Data API. This provides "read only" functions for accessing the seismic sensor-data and

BEAM

meta-data. It is used by the AutoDRM email interface, command line interface, the web interface and for direct program access.

- BDS DataAdd API. This provides the ability to add data to the system. It is used by the Manual and Automated data entry programs.

- BDS Admin API. This provides the ability to add and change the data stored in the system. It is used by local administrators.

The "Database" is used to store Meta Data. This includes information on the sensors, digitisers, the frequency responses, the sensor data files, authorised users database and other information. This database is implemented using MySQL either running on the same platform as the "bdsServer" or on another system.

The Email, Web, Administrator Web interfaces run as separate processes to the main Blacknest Data System Engine (BDSE) "bdsServer". This provides another layer of security against outside attacks as well as providing increased modularity. The Email and Web interfaces could be implemented on other systems if required and other systems can connect directly to the "Data Engine" to get data.

Local programs and utilities are able to access the system using the Data, DataAdd or Admin API's. This allows local network applications such as user analysis programs or robotic data analysis programs to access the data directly.

The Data Converters are implemented as separate modules that can convert to and from their specific data format to a standard internal data API.

Seismic sensor data is stored in an internal data format, BDS. This has been developed for flexibility and the ability to support all of the information in all of the data formats.

The "Data Access Methods" section provides a generic raw data data access interface for the BDSE. This provides the ability to fetch data from different sources using different methods. The primary method is a local file system access to a RAID disk archive, but NFS, FTP, HTTP or other access methods could be added later if required. This module also could be developed to provide caching of the data if required.

The Logging and Statistics module can be used by all other modules to log errors, general information, changes made and statistical information.

There is a set of configuration and other administration tools available to handle configuration and management requirements.

3.1. BDS API's

The BDS API's form the heart of the system. All data and meta data access go through the API's. The BDS provides three separate API's. These will have independent protected access:

- BDS Data API. This provides "read only" functions for accessing the seismic data and meta-data. It is used by the AutoDRM email interface, the web interface and for direct program access.

- BDS DataAdd API. This provides the ability to add data to the system. It is used by the Manual and Automated data entry programs.

- BDS Admin API. This provides the ability to add and change the data stored in the system. It is used by local administrators.

BEAM

The API's are implemented using the BOAP Object Orientated RPC mechanism over a network socket interface. Access to the API's are secured by a userid/password system. This could be extended to use a secure socket interface and perhaps a public/private key system in the future.

The document **BdsApiOverview.html** describes the API in more detail.

The BDS API can be viewed at:
https://portal.beam.ltd.uk/support/blacknest/files/bds/doc/bdsApi/html/index.html

There are example programs in /usr/bds/bdsExamples.

4. The BDS Server

The BDS server is the heart of the system. On this platform runs the main **bdsServer** program. The **bdsServer** program provides the BDS network API for all BDS client applications and implements the core of the system. It provides access to the seismic sensor-data and the associated meta-data. It is responsible for ensuring data security and validating data requests. On this system also runs the BOAP **boapns** daemon. The **boapns** daemon provides an Object name to IP Address and Service number lookup system that the client applications use to find the bdsServer's network API sockets.

The BDS Server will also normally run the MySql database system, although this can be situated on a separate host if required. The raw Seismic data files can also be situated on this host or on a separate system.

The **bdsServer** program has a single configuration file, */etc/bdsServer.conf,* that defines its main configuration settings. It is started automatically at system power up as the **bdsServer** service.

5. The BDS Autodrm

The **bdsAutodrm** program implements the BDS Autodrm email interface. It can be situated on any host that has access to the BDS Server. It implements the Blacknest Autodrm protocol to retrieve seismic data through an email and FTP interface. All data and meta data requests go through the main **bdsServer**. The **bdsServer** is also responsible for converting the data into the requested format. Thus the **bdsAutodrm** is solely responsible fro providing the email interface.

It has a configuration file, */etc/bdsAutodrm.conf*.

6. The BDS Web Interface

The BDS Web interface is implemented as as a PHP web application. It has been based on the BEAM WebSys PHP simple content management system. There is a single PHP module, named bdsdata, that provides the special BDS data access functionality.

The BDS Web interface can be installed on any system that supports an Apache web server and which has access to the BDS Server.

The BDS Web interface is just a demo system at the moment. It makes use of the **bdsDataAccess** command line client program to access the BDS Server. It can search for the data and return sections of the data in any of the BDS supported formats. It also has a primitive graphical data viewer built in.

There is a webserver running on the BdsServer. You can access the BDS web interface using the /bds path. So a URL of the form http://localhost/bds will access the BDS web interface. This is very

BEAM

rudimentary and is there only as an example.

7. The BDS Client Utility Programs

The BDS system has number of command line and GUI applications available to perform administrative functions as well as add data to the system. These use the BDS API's to access the information. The initial applications that have currently been developed include:

1. **bdsAdminGui**. This a simple GUI application that allow basic data and metadata information to be manipulated. It allows all of the available seismic-data and meta-data to be viewed. The meta-data can be edited and verified. Seismic sensor-data can be viewed in graphical form.

2. **bdsDataAccess**. This is a simple command line program that allows access to the seismic data and associated meta data.

3. **bdsImportBlacknestDatabase**. This program is able to import all of the existing MySql instrument response database into the BDS system. It uses the BDS API in order to do this.

4. **bdsImportData**. This command line program is able to import all of the existing seismic data files into the BDS system

5. **bdsImportTapeDigitiser**: This command line program is designed specificity to import the TapeDigitiser analogue sample data files into the BDS system. It first pre-processes the TapeDigitiser data files to re-calculate the block timestamps using the latest VELA timecode decoder. Once this is done the complete set of sensor-data and meta-data from the TapeDigitiser files is imported into the BDS system.

6. **bdsImportScream:** This daemon program reads the incoming GCF files from the Gurlap Satellite data feed and imports the data into the BDS system.

7. **BdsControl:** This program provides access for miscellaneous control operations. It is mainly used for taking a data snapshot for backup.

Future client applications may include:

8. **bdsAddDataGui**. This GUI application would add seismic data to the system. It would present a GUI interface to check the data prior to importing into the BDS system.

9. **bdsAddMetaData**. This GUI applications would allow meta-data to be easily added to the system.

10. **bdsTestGui**. This application could perform sensor-data and meta-data consistency tests.

11. **bdsDataViewGui**. This GUI application would enable the viewing and saving of seismic data from the BDS system.

7.1. BdsAdminGui

This a simple GUI application that allows the sensor-data and meta-data information to be manipulated. It allows all of the available sensor-data and meta-data to be viewed. The meta-data can be edited and verified. Seismic data can be viewed in graphical form. To run use the command "bdsAdminGui -host localhost -user <user>:<password>".

BEAM

7.2. BdsDataAccess

This is a simple command line program that allows access to the seismic data and associated meta data. Use the command without arguments to get a basic usage description. To download some data in BKNAS form you can use a command like:

bdsDataAccess -host localhost -user <user>:<password> -select BN:EKA -startTime 2008-01-04T03:00:00 -endTime 2008-01-04T03:01:00 -command dataGetFormatted -format BKNAS -o data.bknas

7.3. BdsImportBlacknestDatabase

This program is able to import all of the existing instrument response database into the BDS system. This has already be used so it is not that useful for testing unless a complete meta-data import is required.

7.4. BdsImportData

This command line program is able to import all of the existing seismic data files into the BDS system.

1. Data can be imported using the bdsImportData command line program. This can handle any data format supported by the BDS system. It includes: BDRS, GCF and TapeDigitiser data. There is a separate TapeDigitiser data import program that will re-process the TapeDigitisers VELA timecode track and aslo import more of the meta-information from the TapeDigitiser jobInfo files.

2. The user will need to supply the file name, the data format and the network, station and channel names for each of the files data channels. (We could automate the channel names by either using a simple file database or adding a table to the BDS SQL. This would use the time and station name to decide the orientation of channels with the files).

3. The bdsImportData program will first validate the data. The following validations will be performed:

   ● Generally check for data file corruption. This looks at block headers, makes sure that time stamps are in time order and in range.

   ● Check that the blocks time-stamps are continuous. If they are not there are missing blocks of data. This error can be ignored.

   ● Check that the date and time encoded in the file name matches with the time stamps within the data blocks. This error can be ignored.

   ● Check that the sample rate is correct.

   ● Check that there is meta-data information in the BDS system for the channels being imported. This error can be ignored.

   The "-ignore*" flags of the bdsImportData program allows certain validation failures to be allowed.

4. If the validation passes the data is uploaded into the BDS system. If it fails an error message is reported to the user.

There are some example scripts of its usage in /usr/bds/import directory. This code is quite rudimentary.

BEAM

7.5. *BdsImportTapeDigitiser*

This command line program is designed specificity to import the TapeDigitiser analogue sample data files into the BDS system. It first pre-processes the TapeDigitiser data files to re-calculate the block timestamps using the latest VELA timecode decoder. Once this is done the complete set of sensor-data and meta-data from the TapeDigitiser files is imported into the BDS system.

7.6. *BdsImportScream*

This daemon program reads the incoming GCF files from the Gurlap Satellite data feed and imports the data into the BDS system.

8. BDS Data Handling

Seismic data is available in many formats. In order to simplify data access the BDS employs an API that generalises access to the data. The BDS uses the following model for data access.

- All seismic data is stored in individual channels.

- Each channel provides a single set of data organised as a set of timestamped, variable length blocks.

- Each data block has a start and end time stamp and a set of continuous data samples.

- A data block can contain any number of channels of data for the given time period. There can be a variable number of samples per channel.

- Consecutive data blocks may not be continuous in time, there can be missing data blocks. (There could be an option to introduce 0 valued blocks in this case as an option when reading the data).

- The format of each sample is retained through the system. The current Sample types supported include: Int16, Int32 and Float32. When passing the data within the system the samples may be stored in a Float64 for simplicity, but their original sample format is known and can always be re-created without loss. Data is stored on disk in the standard BDS format using the original data's sample format.

- In data files or streams, the samples within a block are packaged in specific pack format. This can employ compression.

- In data files or streams, multiple channels can be multiplexed by sample or by channel. If the data is multiplexed by sample the channels must have been synchronously sampled.

- Meta data for each channel provides, at a minimum: network, station name, channel name, sampleRate and sampleFormat.

- A set of synchronously sampled channel data can be stored in a file and can be marked as synchronously sampled channels. This set of channel data can be output in data formats that require sample multiplexed data such as BKNAS.

The BDS system retains the block sizes and thus time stamps of the original data source as well as the original data's sample format. Thus the full integrity of the data is retained.

BEAM

8.1. BDS Data Selection

In order to access some data in the BDS system it needs to be selected. The BDS has a data selection scheme that allows multiple channels of data to be selected based on regular expressions. The BDS API provides a dataSearch call that takes a set of selection criteria. The selection criteria includes:

StartTime	The start time of the data
EndTime	The end time of the data
Channel Selectors	The set of channels required

Each channel selector has the following parameters:

Network	The network the data is from
Array/Station	The individual station of array of stations that the data is from
Channel name	The channel the data is from
Source	The source of the data.

All of these parameters, except the time fields, can be set to a regular expression. A blank entry is synonymous with "any".

The selection will return information on the data channels available in a DataInfo object. This describes multiple channels of data each being split into multiple time segments. The channels are ordered by network, then source, then station, then finally by channel name. A portion of these data channels can then be selected for data retrieval.

8.1.1. BDS Sources

The Source parameter, used in the BDS system, defines the sensor data's source. For a particular Network/Station/Channel there may be more than one data source over the same time period. For example there may be two digitisers connected to the same set of sensors with different sample rates and/or responses. Also there may be post processed data available. The Source field allows the particular sensor data set and its associated MetaData to be specified. The BDS has a table of all of the possible source values together with a, short form, alias to be used when outputting data in a data format that does not provide the ability to describe the data's source.

The Autodrm system does not, as yet, provide a method of selecting the particular source of data required. Also for a particular period of time one data source should be used, by default, in preference to another due to its better quality or for other reasons. The BDS system provides a Source Priorities table to define which data source should be used over another for a particular period of time if the "Default" keyword is used as the source selector.

The source field in the Selection information describes the data source to be used. If the source is not specified, a null string, then all sources will be selected. If the special keyword "Default" is used, then the source selected will be based on the actual seismic sensor sources available in the chosen time period and the priority order of source selections as listed in the Sources Priority table.

For data where there are no special sources the source name "Main" is normally used.

BEAM

8.2. BDS Seismic Data Files

The system currently supports the following external data file formats:

- BKNAS 1.0 (BKNAS)
- IMS 2.0 (IMS)
- ASCII – simple ASCII data format

The system currently supports import of the following data file formats:

- BDRS
- GCF: Guralp compressed format.
- TapeDigitiser: The Blacknest TapeDigitser file format for Digitised Analogue Tapes.

Later systems will support the following data formats:

- SEED Standard for the Exchange of Earthquake Data format. (Initial support for Blacknest stored SEED files).
- WRA: WRA40, WRA64 and WRA-AGSO
- CD-1.1 IDC Streaming data format

The system stores the seismic sensor-data using a single file format. This is named the **BDS** format. Some more information on this is in the **BdsDataFile.html** document. This format is subject to change as the system evolves.

8.3. BDS Supported File Formats

The BDS System supports the following file formats:

Format	Extension	Description
BDS, BDS-SM	.bds	BDS Internal sample multiplexed file format.
BDS-CM	.bds	BDS channel multiplexed format
BKNAS, BKNAS-1.0	.bknas1	Blacknest AutoDRM data format
IMS, IMS-2.0-CM6	.ims	International Data Centre format, compressed
IMS-2.0-INT	.ims	International Data Centre format, non compressed
GCF	.gcf	Guralp compressed format
TapeDigitiser	.bs, .tdi	Beam, TapeDigitiser data file format
BDRS		Simple, sample multiplexed, binary data format used within Blacknest

BEAM

ASCII		Simple ASCII file format

8.4. Importing Data

Seismic sensor data and meta data can be imported into the system from various BDS client programs. Ideally the meta data for all data channels should be imported before the actual sensor data as this allows more data verification checks to be performed. However, this is not required.

8.5. Importing Meta Data

On a virgin BDS system the meta data can be imported using a suitable program that can import the data from an existing database. For the existing Blacknest MySQL Instrument Response Database there is the bdsImportBlacknestDatabase program. This will go through the existing Instrument Response Database importing all of the information through the BDS API.

Following initial import either the bdsAdminGui or a user written program can import the meta data using the BDS API.

8.6. Importing Sensor Data

There are, currently, three command line programs to import Seismic sensor data. These are:

- bdsImportData: Generic import from data files.
- bdsImportTapeDigitiser: Special importer for TapeDigitiser data files
- bdsImportScream: A daemon that takes the Gurlp SCREAM Satellite data feed and continuously imports the data.

All of these programs are similar in operation. They each require a file name or set of file names to import and a set of information on the data channels. Each import operation will create a single file in BDS format on the BdsServer. It is possible to import a set of sensor data files into one BDS file. This is useful if it is known that all of the channels have been synchronously sampled. In this case the contents of BDS data file can be marked as synchronously sampled and the data can then later be exported in formats that require synchronously sampled data, such as BKNAS.

9. Security: Users and Groups

The BDS system has a basic user account system with user name and password login security. All users and programs are required to log into the system.

9.1. Groups

Each user can belong to one or more groups. The groups define the data and actions that the user can access and/or perform. The groups are configurable and stored in the database. The following main groups are defined:

Group	*Description*

BEAM

admin	General administration functions, read, add, update
control	Control functions: modeSet and snapshotPause
dataAdd	Ability to add data
dataDelete	Ability to delete data
userAdmin	User management
userSet	User or program can set the user
data.Blacknest	Data that only Blacknest staff can access
data.<special1>	Various data only accessible to particular groups

9.2. Data Access Groups

There are a set of data access groups. These define particular sets of sensor-data and meta-data that require the user to belong to a particular group in order to be able to access. Each AccessGroup entry has a period of time, a Network and an Array/Station description that is associated with a particular group.

10. BDS Notes system

The BDS notes systems allows users and programs to add notes into the BDS database. These can be generic or related to particular data.

11. BDS Changes System

The BDS system has a meta-data change logging system. This allows all changes made to the BDS to be logged. The changes system is organised into Change Groups. Whenever a user logs into the system a new Change Group is added with information on the user and what the reason for access is. Change groups can be added at any time whenever a set of work is going to be carried out.

Each Change Group has then a list of database changes that occurred listing the table name and the record id.

The Changes information can grow large quite quickly and so there is administrative functions to clean up the changes information.

12. BDS DataBase

The BDS system employs a MySQL database to store Seismic MetaData information such as Instrument responses and information on the available Seismic data. It also stores information on the users and is used for queuing user requests for data.

There is a single database, named **BDS**, that has a number of tables. The **BdsSchema.html** document will describe the prototype Database Schema in detail.

Normally only the **bdsServer** program accesses the database although it obviously is possible to directly access it. The **bdsServer** also provides an API call to directly access the database if required.

BEAM

13. Notes

1. Accessing a large time period of data in the Web interface will cause an out of memory exception in the PHP web interface.

2. There is no high level commands to allow user orientated meta-data updates as yet. At the moment the bdsAdminGui program only allows for manual low level database field updates. Higher level functionality such as "Add this set of channels with settings for this time range ..." would be good. Blacknest will need to make a list of the main higher level functions that would be initially implemented. It is envisage that a simple scripting language would be used to implement these changes.

3. There is no support for Leap seconds.

4. The BdsWeb interface is basic and only an example at the moment.

5. How to select when there is data overlap ? If data from the TapeDigitiser it would be best to select the later data file as the change may have happened due to head cleaning.

6. System requires Sample rate to be the same on all channels if data is asked for in a synchronously sampled block form. We could re-sample to the highest sample rate if required, with appropriate warning.

7. What if there are data holes within a time period ? Should we only allow data sets to be complete or would it be best to output 0's in the holes with appropriate warning messages ?

8. What happens when meta-data changes with the time period ? At the moment this is ignored. We could add certain meta data to the output data blocks and output blocks synchronised to the meta data change.

9. Need to add support for leap seconds. Probably best to have a flag in each data block which is set if a leap second change occurred within the block. The number of samples would not match the time stamp period by 1.

13.1. Existing Data Notes

During import of existing data and checking of BKNAS and IMS data output formats we have noticed the following possible data issues with the current data:

1. Some of the Autodrm old database's floating point values seem to be truncated. This includes Latitude, Longitude, elevation etc.

2. Some of the Autodrm old database's time stamps look a little odd.

3. Quite a lot of the EKA and WOL data fails to import with time stamp validation errors. I have not looked into this as yet.

4. We have seen EKA BDRS files with a first block having timestamps that appear to be incorrect.