

Advanced data extraction infrastructure: Web based system for management of time series data

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2010 J. Phys.: Conf. Ser. 219 042034

(<http://iopscience.iop.org/1742-6596/219/4/042034>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 212.42.192.24

This content was downloaded on 22/01/2014 at 08:11

Please note that [terms and conditions apply](#).

Advanced Data Extraction Infrastructure: Web Based System for Management of Time Series Data

S Chilingaryan¹, A Beglarian¹, A Kopmann², S Vöcking²

¹ Forschungszentrum Karlsruhe, Hermann-von-Helmholtz-Platz 1, 76344
Eggenstein-Leopoldshafen, Germany

² University of Münster, Institut für Kernphysik, Wilhelm-Klemm-Strasse 9, 48149 Mnster,
Germany

E-mail: Suren.Chilingaryan@kit.edu

Abstract. During operation of high energy physics experiments a big amount of slow control data is recorded. It is necessary to examine all collected data checking the integrity and validity of measurements. With growing maturity of AJAX technologies it becomes possible to construct sophisticated interfaces using web technologies only.

Our solution for handling time series, generally slow control data, has a modular architecture: backend system for data analysis and preparation, a web service interface for data access and a fast AJAX web display. In order to provide fast interactive access the time series are aggregated over time slices of few predefined lengths. The aggregated values are stored in the temporary caching database and, then, are used to create generalizing data plots. These plots may include indication of data quality and are generated within few hundreds of milliseconds even if very high data rates are involved. The extensible export subsystem provides data in multiple formats including CSV, Excel, ROOT, and TDMS. The search engine can be used to find periods of time where indications of selected sensors are falling into the specified ranges. Utilization of the caching database allows performing most of such lookups within a second. Based on this functionality a web interface facilitating fast (Google-maps style) navigation through the data has been implemented.

The solution is at the moment used by several slow control systems at Test Facility for Fusion Magnets (TOSKA) and Karlsruhe Tritium Neutrino (KATRIN).

1. Introduction

Advanced Data Extraction Infrastructure (ADEI) has been developed to provide data exploration capabilities to a broad range of physical experiments dealing with time series. Examples of such systems include slow control systems maintaining archive of sensor registrations for long periods of time, cosmic ray experiments evaluating fluxes of cosmic rays incident on the detector surface, meteorological systems measuring various aspects of weather, etc. All these systems have very different characteristics: amount of data channels, their types, sampling rates, etc. The data acquisition and control subsystems are storing data in many different ways utilizing various data formats and underlying database engines. On the other side users need information in different data formats which are supported by analysis tools they use for post processing. Besides, operators need a tool providing possibility to examine all collected data checking the integrity and validity of measurements. It also needed to search and export data possessing specified characteristics.

To provide such broad coverage ADEI utilizes highly modular architecture. The system consists of backend and frontend parts communicating over HTTP protocol using Asynchronous JavaScript and XML (AJAX [1]) approach. The ADEI backend defines few abstract interfaces which are used to implement various capabilities using simple plugins. The data sources are interfaced with dedicated drivers implemented data access abstraction layer. The higher levels of system are relying on this abstract interface to get data in a uniform way from arbitrary storage.

The ADEI web frontend is inspired by *GoogleMaps* interface. The navigation through the data is feasible using mouse only. Single or multiple time series are plotted using the data from currently selected time interval. Then, the plot could be dragged and zoomed over time and value axes. Detailed information is provided about the graphics around mouse pointer. The region of plot may be selected and exported in one of supported formats.

ADEI is licensed under GNU General Public License and uses only free open source technologies. It is currently implemented at slow control systems of Karlsruhe Tritium Neutrino (KATRIN [2]) experiment and Test Facility for Fusion Magnets (TOSKA). There is also experimental implementation providing access to the data collected at Aragats Space Environmental Center (ASEC [3, 4]).

The rest of article is organized as follows. The second section briefly reviews ADEI architecture and supported applications. The third section describes features of ADEI web frontend.

2. Architecture

ADEI is designed to deal with the data sampled at high rates and stored for long periods of time. ADEI implementation at TOSKA deals with the data continuously sampled at 10 Hz. The weather tower at Forschungszentrum Karlsruhe has archive of measurements since beginning of 20th century. Processing such amounts of data requires enormous computational power. However, the interactive tools should operate in near real-time and extract important information from this enormous amount of data. To achieve this goal ADEI continuously monitors incoming data, performs preprocessing and caches important information in a high performance database.

The simplified diagram of ADEI architecture is presented on Figure 1. The main logic of ADEI system is contained in a backend which is implemented purely in PHP programming language. The backend incorporates a data access layer, a caching daemon, an ADEI library. Communication with the web frontend and other client applications is maintained using web services. HTTP protocol is used for data exchange, XML and JSON - for data encoding.

2.1. Backend

The data access layer hides details of underlying data sources providing other components of the system with a uniform way of data access. The data is organized hierarchically. The top level of hierarchy is the data source and ADEI may underline several data sources. The time series provided by the data source is divided in time-synchronized groups, so called *LogGroups*. The current version includes drivers to access data stored in:

- (i) Relational databases accessible using PDO or ODBC modules of PHP. Most of popular databases including MySQL, PostgreSQL, Oracle, and Microsoft SQL server are supported.
- (ii) ZEUS (The Central Data Acquisition and Control System [5]) control system
- (iii) ASEC data format
- (iv) RRD (Round Robin Database Tool [6]) data format

The caching daemon is continuously running on a backend server and polls all data sources for a new data. When the data is acquired it piped through series of filters which check the

Applications

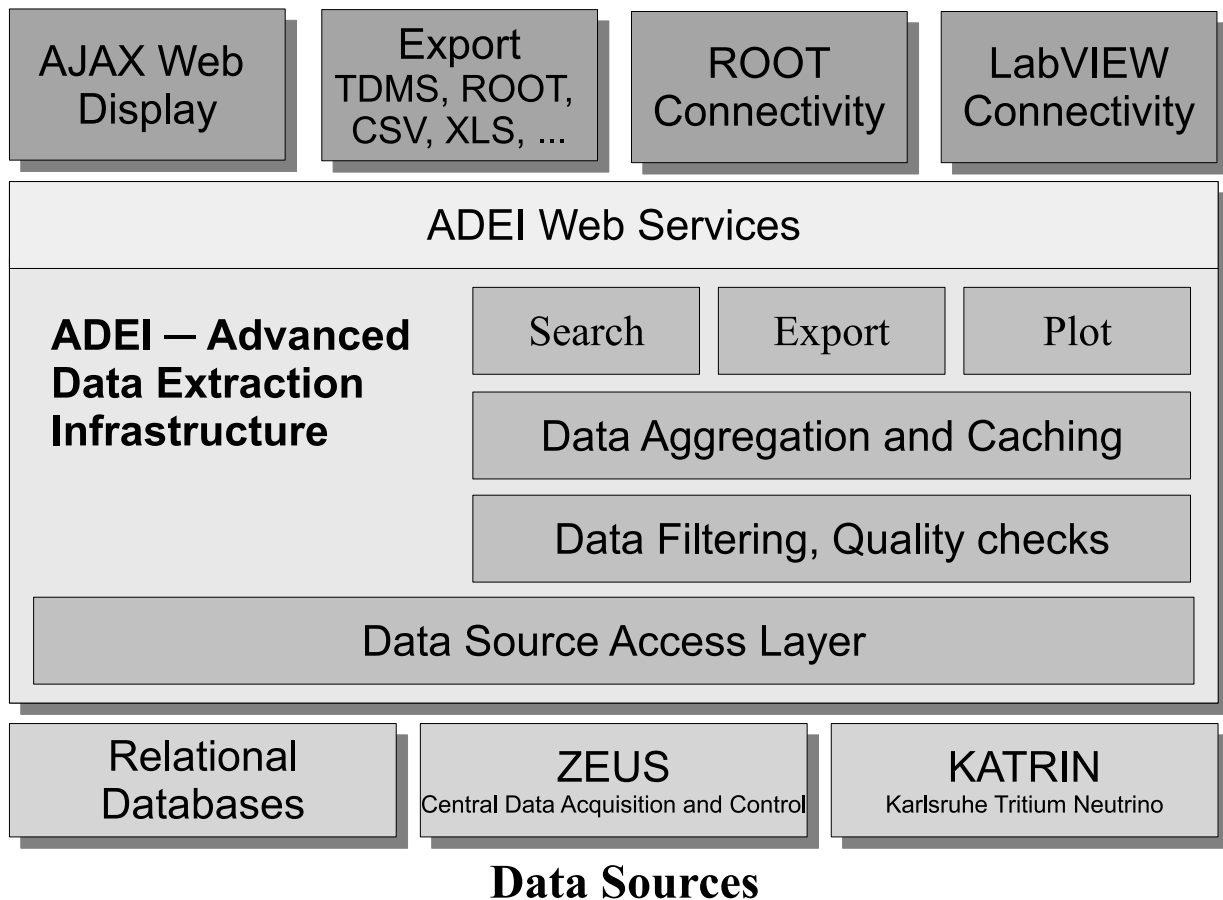


Figure 1. Architecture of Advanced Data Extraction Infrastructure. Data Source Access Layer unifies access to the time series stored in different formats. After data filtering and quality checks the data is aggregated and stored in intermediate caching database. Access to the data is provided by ADEI library and web services are used to communicate with client applications.

data quality, apply correction coefficients and drop invalid data. Then, the data is aggregated over intervals of few different sizes. For each size, called *cache level*, five new time series is constructed and stored in MySQL database (*caching database*). This time series contain following information for each interval of aggregation:

- (i) Number of records
- (ii) Longest time interval without valid data records
- (iii) Average value
- (iv) Minimum value
- (v) Maximum value

The ADEI PHP library implements several classes which are using the caching database and data access layer to provide stored data in various forms. The plot module first selects the maximal cache level providing enough points to generate the plot of specified size. The configuration of caching database is selected in the way what for an arbitrary time interval it is

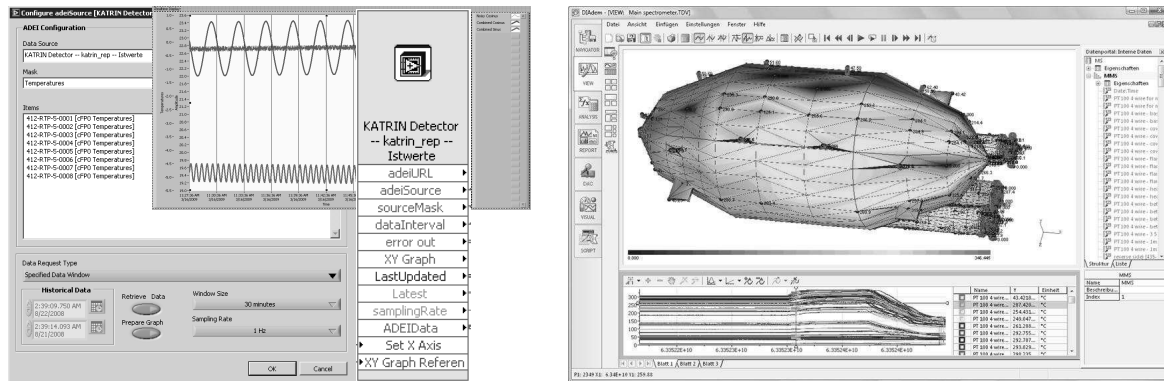


Figure 2. ADEI Express VI enables ADEI support in LabVIEW based control applications (left). During commissioning of the KATRIN main spectrometer a 3D temperature monitor has used ADEI interface to access slow control data (right).

possible to find appropriate cache level providing between 1000 and 10000 data records within the specified interval. Such amount of points fulfills most of plotting demands and in the same time the plots could be generated relatively fast. After selection of cache level is made, the data is extracted from correspondent caching tables and one of supported algorithms is used to convert aggregated values into the graphic points. The JpGraph is used to render a plot [7].

2.2. Fronted

ADEI frontend is implemented using AJAX approach. The pages are rarely reloaded, required information is obtained from the ADEI backend using XMLHttpRequest object, and JavaScript is used to update content of the page [1,8]. The data plots are generated by the backend and delivered to frontend as PNG images.

The open source JavaScript libraries are widely used to implement advanced features.

- (i) *Prototype* framework is used to transparently support multiple browsers [9].
- (ii) *dhtmlHistory* is providing navigation history and bookmarking support in AJAX context [10].
- (iii) *DHTMLX* components are providing menu and source tree objects [11].
- (iv) *ImageCropper* is modified and used to select subregions of plot [12].
- (v) *YUI Compressor* is used to compress JavaScript code [13].

2.3. Client Applications

The ROOT analysis framework and National Instruments LabVIEW are extended with ADEI support. The C++/ROOT library is interfacing all features of ADEI web services and used in KATRIN database project. An Express VI is simplifying access to ADEI data for National Instruments LabVIEW applications. Currently, it is exploited by a real-time monitor for TOSKA slow control system and KATRIN 3D temperature monitor used during KATRIN main spectrometer commissioning to visualize tank heating. See Figure 2 for illustrations.

3. ADEI Web Interface

The main view of ADEI web frontend is represented on Figure 3. The screenshot is taken using real system running at KATRIN. It incorporates data from the KATRIN slow control system and few simulated sinusoidal signals. Numeric labels from 1 to 12 are used to reference interface elements in the description below.

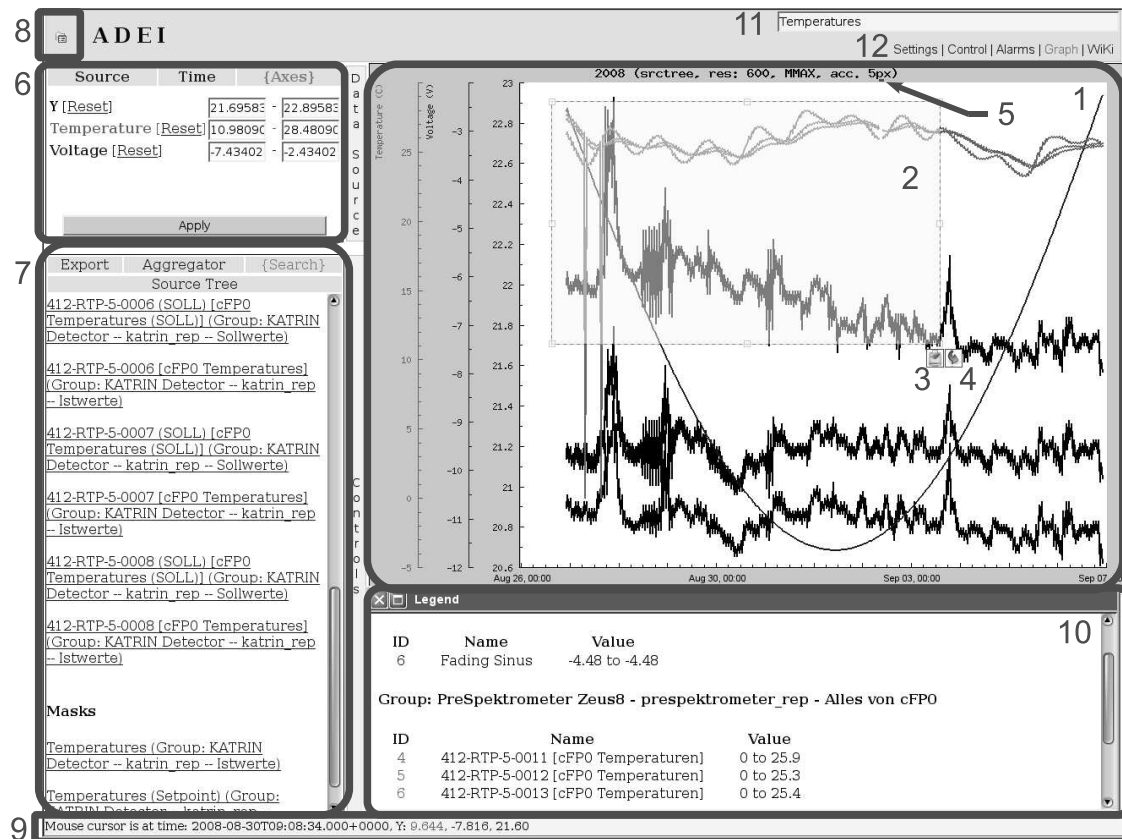


Figure 3. Screenshot of ADEI Web Frontend. The data from KATRIN slow control system is rendered in the plot. The data outage is indicated using a small line on top of the plot (see 5). Legend contains description of displayed graphics. The selected part of plot may be zoomed or exported using buttons 3 and 4. Axes controls and results of search are located in the left sidebar.

The main window (*label 1*) contains plot depicting measurements of 7 sensors over two week period from August 26, 2008 to September 07, 2008. The registrations of sensors were sampled into the database approximately ten times in a second what gives about 12 millions of data records over two week interval. To handle such amount of data, ADEI aggregates data and produces approximately few thousand data points which are, then, used to plot a graph.

The overall aggregation and plotting performance is near to real-time. ADEI intensively uses caching techniques to accelerate data aggregation. The actual performance is highly depends on the computational power of server hardware and its I/O throughput. However, on the most of up to date systems the complete time of plot generation does not exceed 500 milliseconds.

ADEI supports multiple value-axes on a single plot. The amount of supported axes is only limited by the size of browser window (any single axis needs approximately 75 pixels). The axes information is normally provided by the data source and can be overridden in ADEI configuration. The color encoding is used to show correspondence between axes and graphics on the plot: few primary colors are used to represent axes and their undertones are used for each of the graphics. In the presented example, totally three axes are used. One is used for a single voltage sensor, another for three temperature sensors, and a last one depicts three remaining sensors for which axis information is not available.

The rest of section describes ADEI user interface and its subsystems.

3.1. Data Navigation

The subarea of plot can be selected using mouse pointer while holding left button (*label 2*). After selection is made it still can be fine tuned: resized or positioned using mouse or keyboard arrows. The buttons in the right-bottom part are used to export data within selected time interval (*label 3*) or to zoom into the selection (*label 4*). Additional functional buttons can be implemented using custom plugins. For example, ADEI installation at KATRIN uses a special button to search for experiment runs within selected time interval. If runs are found the run list along with descriptions is displayed in the sidebar (*label 7*). Version designed for FZK weather tower provides an extra button to mark selected data as invalid.

The data navigation is following *Google-maps* standards. The current plot on display can be zoomed in and out by scrolling mouse wheel. The default action is to zoom along time axis at the position of mouse pointer. However, the key modifiers may be used to zoom over value axis or zoom in the center of the plot. The adjustments of plot position on the time and value axes are achievable by scrolling mouse over the correspondent axis. The support for dragging plots along axes is included in development branch. The double click on considered axis will restore it into the automatic mode and all available data will be displayed.

The time series on display are selected and adjusted using several methods:

- (i) The *Source* tab of the upper sidebar (*label 6*) contains a selectable list of channel groups. These groups are provided by data source or/and set in the ADEI configuration. It is possible to plot a whole set of channels from the selected group or any single item from it.
- (ii) The *Source Tree* tab of the bottom sidebar (*label 7*) contains hierarchical tree of all time series available in the system. It is possible to select and plot any desired subset of them.
- (iii) Fast switching between the data channels and channel groups on display is possible by scrolling mouse wheel while holding the correspondent modifier key.

ADEI supports navigation history. *Forward* and *Back* buttons of the browser could be used to go back and forth in the history. The URL in the navigation bar is always precisely describing current position, selected time series, and all configured properties. This URL could be sent to the colleagues over e-mail and exactly the same plot will be displayed on their PC.

3.2. Plot Interaction

A status bar (*label 10*) is used to provide status and contextual messages to the user. Currently performed actions, their completion status, contextual help, emerging error messages are reported using status bar. On mouse movement the position of mouse pointer along all axes is reported as well. An example could be seen on the provided screenshot. The color coding is used to help with association of axes. On a key presses the help message describing effects of this key when used as modifier while scrolling mouse wheel, dragging its pointer or pressing button is displayed.

In order to simplify analysis of complex plots with big amount of curves, ADEI provides possibility to investigate graphics passing in the specified area of the plot. A legend window (*label 10*) is popped up when the left mouse button is clicked. It contains a list of all graphics on the plot which are passing near position where mouse was clicked. The short name, description, and a range of values possessed in the neighborhood are presented on the legend. Current development branch allows selecting subset of channels from the legend. These channels may be removed from the plot or displayed solely. The color coding is used to associate graphics on the plot with information in the legend.

3.3. Search Capabilities

ADEI includes an extensible search engine. It support pluggable search modules and able to search for channels, channel groups, channel values, and time intervals.

The channel search is very flexible. The channel names and descriptions are searched for single and multiple words, exact phrases, and regular expressions. The words are matched in three different ways:

- (i) The words should match exactly
- (ii) The matched word need to start from the search term but may have arbitrary suffix
- (iii) The matched word needs to contain the searched term

The search is case-insensitive and all these types of matching can be mixed in a single search query.

The value search finds a set of time intervals where the values of the given channel are above/below the specified threshold. The two modes are supported: search for time intervals where any value from the interval is above/below the specified threshold and search for time intervals where at least some of the values are above/below the threshold. The data cache is used to accelerate searches over big amounts of data and, therefore, precision of found intervals is depending on the caching configuration. The current version of ADEI is only supporting searches on a single data channel.

The interval search allows users to quickly position time axes. The search module supports strings like *January 2005* or *January - March, 2006* and upon submitting of search request the time axis are set accordingly.

The search input (*label 11*) is located in the right-top corner of browser window. Upon search completion, the bottom sidebar (*label 7*) is opened and results are reported in the *Search* tab. The example presented on screenshot displays results of searching for temperature sensors in the KATRIN slow control database.

3.4. Missing and Invalid Data

Often due to the sensor failures, power outages, and other problems in the control and data acquisition systems, the collected information includes periods when no data was recorded or existing recordings are invalid. Normally, operators should be able to quickly locate problematic intervals and investigate them. However, using aggregating plots over lasting time intervals it is rather difficult to notice outages of short duration. On the yearly surveys the day data is represented by a few points only. It makes impossible to visualize problematic intervals shorter than a quantity of hours and in the same time appropriately plot the valid data.

In order to handle such situations, ADEI includes a quality indication line on a top of the data plot (just below a plot title). On the screenshot it is possible to see a tiny line indicating short period when the data was not recorded due to restart of the slow control system (see *label 5*).

3.5. Data Export

ADEI supports multiple export formats. The current version includes support for CSV (Comma Separated Values), Microsoft Excel, ROOT (an analysis framework for high energy physics [14]), and TDMS (Technical Data Management Streaming [15]). The separator and decimal points used in CSV output are specified in the ADEI configuration to support users with different locale settings.

Additional formats may be implemented in two ways. The ADEI support custom export plugins. Alternatively, it is possible to filter exported data using system scripts. For example, to generate ROOT output, the data in CSV format is piped to standard input of a simple ROOT application which converts it to ROOT format and prints to standard output. The same mechanism could be used to compress output before returning it to the client application. The chains of filters are supported. This allows producing of archived ROOT files.

To limit amount of exported data, a data resampling may be requested by user. When switched on, the export subsystem will skip some of the data records to produce data at frequency lesser than the specified rate.

The properties of export subsystem are configured in the *Export* tab of sidebar (*label 7*). Export button extracts all data currently on display. A part of data may be exported using area selection (*label 3*).

3.6. Sidebar and Menu

ADEI sidebar provides fine grained control over the ADEI behavior. To save plotting space the sidebar may be hidden when not used. It consists of two parts. The top part (*label 6*) includes three tabs controlling the data displayed on the plot: time series and axes configuration.

- (i) *Source* tab provides controls selecting channels and channel groups to be displayed.
- (ii) *Time* tab contains configuration of time axis. Few different modes are supported. The data source may provide a list of time intervals when something important was happening. It is possible to select a desired interval from this list and apply it to the time axis. Alternatively, the beginning and the end of time axis could be set manually with a microsecond precision. Other options include visualization of all stored data or just last quantity of seconds (i.e. plot for last minute, hour, day, week, etc.). In the last case the plot will be periodically updated to display incoming data. In the case if the data recorded in non-persistent fashion or with a high latency it is possible to plot last block of recordings instead of displaying the data for recent time.
- (iii) *Axes* tab allows to fine tune the value axes by specifying their ranges. Only axes currently used on the plot are displayed. The color encoding is used to help with axes association.

The bottom part of sidebar (*label 7*) includes options controlling behavior of ADEI subsystems and could be extended with plugins.

- (i) *Export* tab encompasses export settings: export format, resampling rate, etc.
- (ii) *Aggregator* tab contains aggregation settings. It is possible to configure aggregation algorithm, to switch usage of data interpolation, and to adjust the mode of the quality indicator (see section 3.4).
- (iii) *Source Tree* tab implements hierarchical tree of data channels. The arbitrary group of channels may be selected and plotted.
- (iv) *Search* is a place holder for search results.

3.7. Drop Down Menu

The drop down menu is located in top-left corner of browser window (*label 8*). It collects in a single place all controls otherwise distributed over the user interface. Here it is possible to select channel group and time interval, set aggregation algorithm, adjust export properties, etc.

3.8. Control Subsystem

A simple control subsystem is included in ADEI. It incorporates information about current system behavior and allows setting of the setpoints (if supported by data source). The control channels do not possess historical data, but store only the current values along with two timestamps:

- (i) Timestamp indicating when the current value was set
- (ii) Timestamp indicating when it was last time verified

ADEI Settings | Control | Alarms | Graph | Wiki

List of alarms for: KATRIN Detector - katrin_rep

Severity	Alarm	First Seen	Last Seen	Count	Description
0	Valve 2 off	29 Jul 2008, 09:21:57	12 May 2009, 17:31:01	3	
0	Valve 3 off	29 Jul 2008, 09:21:57	12 May 2009, 18:59:18	27	
0	Interlock Status 5 (hardware) ok	01 Aug 2008, 14:27:27		21	
0	Interlock Status 2 (hardware) ok	19 Aug 2008, 19:19:30	12 May 2009, 17:30:32	50	
0	Interlock Status 3 (hardware) ok	19 Aug 2008, 19:19:31	12 May 2009, 18:59:23	64	
0	Interlock Status 4 (hardware) ok	19 Aug 2008, 19:19:32		44	
0	Interlock Status 6 (hardware) ok	19 Aug 2008, 19:19:33		44	
0	Interlock Status 7 (hardware) ok	19 Aug 2008, 19:19:34		43	
0	Interlock Status 8 (hardware) ok	19 Aug 2008, 19:19:34		103	
0	Valve Status 2 off	20 Aug 2008, 07:00:08	12 May 2009, 17:30:32	12	
0	Valve Status 3 off	20 Aug 2008, 07:00:08	12 May 2009, 18:59:19	36	
0	Valve Status 4 off	20 Aug 2008, 07:00:08		10	
0	Valve Status 5 off	20 Aug 2008, 07:00:08		9	
0	Valve Status 6 off	20 Aug 2008, 07:00:08		10	
0	Valve Status 7 off	20 Aug 2008, 07:00:08		10	
0	Valve Status 8 off	20 Aug 2008, 07:00:08		9	
0	Electrical Heater Mask 3 off	20 Aug 2008, 08:22:17		76	

Figure 4. Screenshot of Alarm View. The alarms registered by KATRIN slow control system are listed.

Some data sources may provide list of currently active alarms or alarms occurred during the specified time interval. Provided information include alarm name and description, its severity, the timestamps when it was first seen and last gone, etc. The plot and alarm views are connected. By default the time interval used to render current plot is used to extract alarm information from the data source. This means that in the alarm view ADEI lists all alarms affecting the data currently plotted. A screenshot on Figure 4 displays alarm view.

3.9. ADEI Wiki

In the big systems with a large number of data channels and long data collection intervals it is rather difficult for new users to find important data. It is extremely helpful to have a descriptive listing of important channel groups along with links to the actual plots. The multiple small previews collected at one page may provide indispensable information about current system status.

This functionality is provided by Wiki engine included in ADEI web interface. Besides standard Wiki syntax it includes few ADEI-specific extensions:

- (i) *[preview]* - generates a preview plot. The channel group, time interval, image size, aggregation mode, and other standard properties may be specified. The preview is linked and upon a click will switch to the plotting view and display appropriate graph.
- (ii) *[grouplist]* - generates linked previews for all channel groups available in the system.
- (iii) *[channels_by_name]* - includes alphabetical listing of all channels in the system. Upon a click the selected channel will be plotted.
- (iv) *[channels_by_group]* - includes hierarchical listing of all channels in the system. The selected channel will be plotted upon a click as well.

The example Wiki is depicted on Figure 5. It includes previews of three groups for last 24 hours and alphabetical list of data channels below.

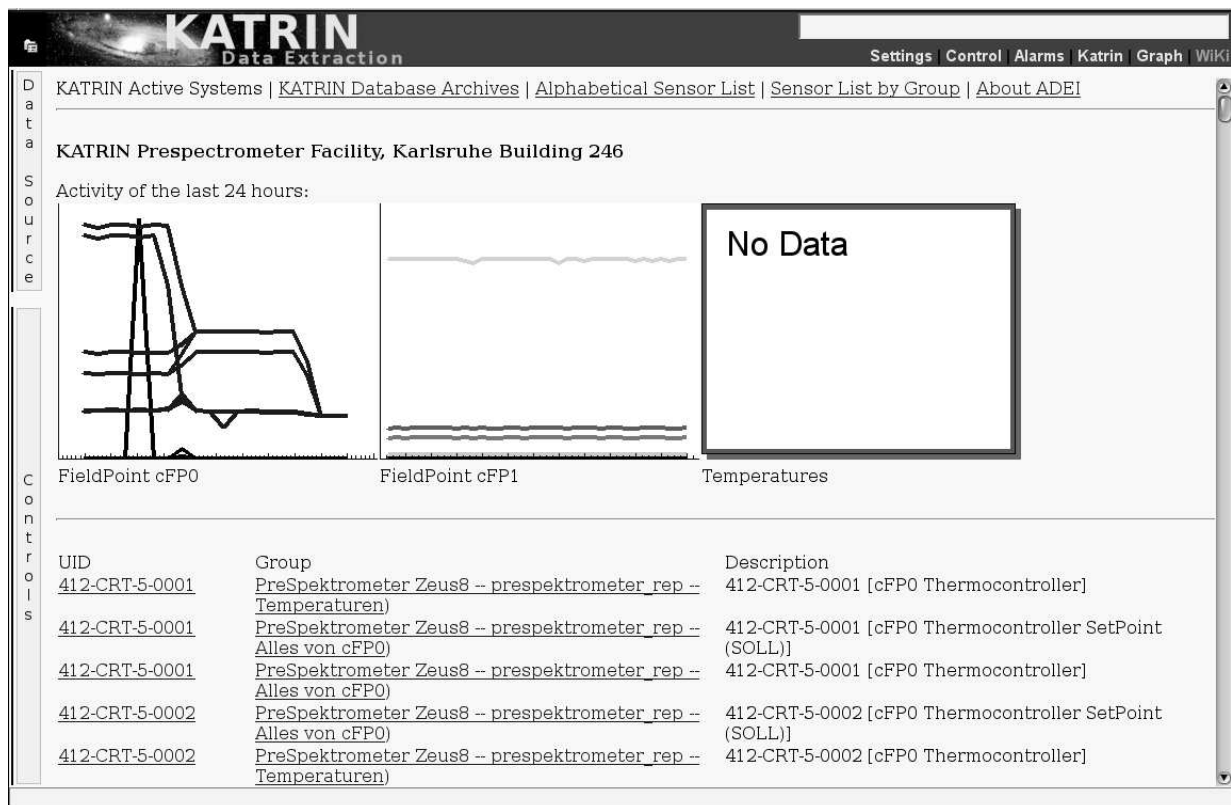


Figure 5. Screenshot of ADEI Wiki designed for KATRIN slow control system. The page displaying activity for last 24 hours is open. Previews are depicting status of three KATRIN subsystems. According to the third preview the temperature database is currently offline. The alphabetical list of sensors follows previews.

References

- [1] Brett McLaughlin 2005 Mastering ajax URL <http://www.ibm.com/developerworks/web/library/wa-ajaxintro1.html>
- [2] KATRIN Collaboration 2005 Katrin design report 2004 Tech. rep. Forschungszentrum Karlsruhe URL <http://www-ik.fzk.de/katrin/publications/documents/DesignReport2004-12Jan2005.pdf>
- [3] Chilingarian A, Avakyan K, Babayan V *et al.* 2003 *J. of Phys. G* **29** 939–951
- [4] Chilingarian A, Arakelyan K, Avakyan K *et al.* 2005 *Nucl. Inst. Meth. D* **A543** 483–496
- [5] Lefhalm C and Krieger V 2005 *Proc. of 14th IEEE-NPSS Real Time Conference* p 132
- [6] RRDtool team 2009 Round robin database tool URL <http://oss.oetiker.ch/rrdtool/>
- [7] Aditus Consulting 2009 Jpgraph - php graph creating library URL <http://www.aditus.nu/jpgraph/>
- [8] W3C 2008 The xmlhttprequest object URL <http://www.w3.org/TR/XMLHttpRequest/>
- [9] Prototype Core Team 2009 Prototype javascript framework URL <http://www.prototypejs.org/>
- [10] Dillard B and Neuberger B 2007 Really simple history (rsh): Ajax history and bookmarking library URL <http://code.google.com/p/reallysimplehistory/>
- [11] DHTMLX 2009 Dhtml extensions - professional ajax components for advanced web ui. URL <http://www.dhtmlx.com/>
- [12] Spurr D 2006 Javascript image cropper ui, using prototype and script.aculo.us URL <http://www.defusion.org.uk/code/javascript-image-cropper-ui-using-prototype-scriptaculous/>
- [13] Yahoo 2009 Yahoo! ui library: Yui compressor URL <http://developer.yahoo.com/yui/compressor/>
- [14] The ROOT Team 2009 Root: A data analysis framework URL <http://root.cern.ch>
- [15] National Instruments Ni tdm data model URL <http://zone.ni.com/devzone/cda/tut/p/id/3727>