

---

Subject: Proposal Statistics

Posted by [markus.ullius](#) on Mon, 15 Nov 2004 19:26:19 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Hello

Since I'm working on a statistics-export for OpenTimeTable I have the following proposals for the STATISTIC-Data:

- 1)change the name of "source" to "dataSource" - as it is in TRAIN
- 2)add an entry "statisticType" having the values "mean", "median", ... to have different statistics
- 3)maybe one could also add the field "type" as in ENTRY having the values "stop", "pass" and so on.
- 4)For standard deviations and so on there would be needed some float values but I have to think about this in detail - it's just a first idea

Example see below

Best Regards

Markus Ullius

```
<train trainID="6203" type="planned" dataSource="opentimetable"
dataSource="planned">
<timetableentries>
<entry posID="WH" departure = "05:34:00" departureDay="0" type="begin">
<statistic source="opentimetable" departure="05:34:46" departureDay="0"
type="begin" statisticType="mean">
</statistic>
<statistic source="opentimetable" departure="05:34:46" departureDay="0"
type="begin" statisticType="median">
</statistic>
</entry>
<entry posID="LZ" arrival = "05:56:00" arrivalDay="0" type="end">
<statistic source="opentimetable" arrival="05:58:03" arrivalDay="0"
type="end" statisticType="mean">
</statistic>
<statistic source="opentimetable" arrival="05:58:03" arrivalDay="0"
type="end" statisticType="median">
</statistic>
</entry>
</timetableentries>
</train>
```

---

Subject: Re: Proposal Statistics

Posted by [Daniel Huerlimann](#) on Tue, 07 Dec 2004 07:25:15 GMT

In article <cnavsm\$4u8\$1@sifa.ivi.fhg.de>,  
markus.ullius@sbb.ch (Markus Ullius) wrote:

```
....  
>  
> <train trainID="6203" type="planned" dataSource="opentimetable"  
>   dataStatus="planned">  
>   <timetableentries>  
>     <entry posID="WH" departure = "05:34:00" departureDay="0" type="begin">  
>       <statistic source="opentimetable" departure="05:34:46" departureDay="0"  
>         type="begin" statisticType="mean">  
>       </statistic>  
>       <statistic source="opentimetable" departure="05:34:46" departureDay="0"  
>         type="begin" statisticType="median">  
>       </statistic>  
>     </entry>  
>     <entry posID="LZ" arrival = "05:56:00" arrivalDay="0" type="end">  
>       <statistic source="opentimetable" arrival="05:58:03" arrivalDay="0"  
>         type="end" statisticType="mean">  
>       </statistic>  
>       <statistic source="opentimetable" arrival="05:58:03" arrivalDay="0"  
>         type="end" statisticType="median">  
>       </statistic>  
>     </entry>  
>   </timetableentries>  
> </train>  
>
```

Hello Markus

Thank you for your recommendations regarding the statistical timetable data. I like the idea to exchange this type of data within RailML.

There is just one question: Shouldn't we have some additional information about the statistical data like the begin and the end date of period where the resulting statistics were measured?

Best regards

Daniel Huerlimann

---

Subject: Re: Proposal Statistics  
Posted by [markus.ullius](#) on Mon, 20 Dec 2004 15:47:56 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Daniel Huerlimann wrote:

- > There is just one question: Shouldn't we have some additional
- > information about the statistical data like the begin and the end date
- > of period where the resulting statistics were measured?

When you also export real timetable data you can determine where the statistics are calculated from. If you only have a start- and an end-date what will happen if there are some missing days in between?

Best Regards  
Markus

---

---

Subject: Re: Proposal Statistics  
Posted by [Joachim.Rubröder](#) on Thu, 27 Jan 2005 12:35:12 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hello Markus,  
thank you for your new inputs about statistics.

If you agree, I'd prefer to delete the two STATISTIC attributes 'source' and 'date' and take the attribute group 'dataReferences' instead (including 'dataSource', 'dataDateTime' and 'dataStatus' as it is used in TRAIN)

The attribute 'type' as in TRAIN can be added for STATISTIC as well, but I'm thinking of a more detailed new element 'stopType' for the TRAIN describing the different kinds of conditional and operational stops. Should such a new element be used for STATISTIC in the same way?

The 'statisticType' seems to be reasonable as well, what values besides "mean" and "median" could be used there?

And what about your attributes needed for standard deviation?

best regards  
Joachim Rubröder

---

---

Subject: Re: Proposal Statistics  
Posted by [markus.ullius](#) on Fri, 04 Feb 2005 10:01:15 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hello Joachim

- > The attribute 'type' as in TRAIN can be added for STATISTIC as well, but

- > I'm thinking of a more detailed new element 'stopType' for the TRAIN
- > describing the different kinds of conditional and operational stops.
- > Should such a new element be used for STATISTIC in the same way?

Maybe this may make sense - also to be 'compatible' with the non-statistical parts.

- > The 'statisticType' seems to be reasonable as well, what values besides
- > "mean" and "median" could be used there?

There could also be some kind of %-mean, %-median, indicating the mean and median for the best x % of the trains -> value-field for the %-value would also be required

- > And what about your attributes needed for standard deviation?

The standard deviation would require some kind of float-value giving the stdDev - whereas the mean and median could either be "timestamps" or "delayvalues" indicating the delay from the "scheduled-timestamp"

Best regards  
Markus Ullius

---

Subject: Re: Proposal Statistics  
Posted by [andreas.voss](#) on Tue, 15 Feb 2005 14:00:42 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Markus Ullius wrote:

- >> There is just one question: Shouldn't we have some additional
- >> information about the statistical data like the begin and the end date
- >> of period where the resulting statistics were measured?
- > When you also export real timetable data you can determine where the
- > statistics are calculated from. If you only have a start- and an end-date
- > what will happen if there are some missing days in between?

Hi Markus

of course, there are many aspects one MIGHT want to take care of when modelling statistical data - to get some ideas, have a look at the work of the SDMX (statistical data and metadata exchange, [www.sdmx.org](http://www.sdmx.org)) initiative. Certainly, not all of this is necessary within the realm of RailML and I am not sure how far we actually need to go.

Maybe, the main concern should not be an in-depth description of the origin of the dataset (dates, methods and tools of analysis etc.), but to

allow a way of modelling in which each timetable entry in a file CAN have a unique identifier (e.g., in order to avoid conflicts when the data is imported into a relational database or an application that is based on one).

For example, if you have the mean departure time of March and the mean departure time of calendar week 11 in the same file, an application that is reading the file must be able to make a distinction. Whether this is based on an somewhat arbitrary identifier ("DatasetMarkus418b" or "MonthlyStatisticsMarch2005") or an explicit listing of all parameters describing the dataset and its origin is - in my opinion - of minor concern.

Regards

Andreas

---

---

Subject: Re: Proposal Statistics  
Posted by [markus.ullius](#) on Fri, 25 Feb 2005 09:10:03 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

I think it's a good idea to have a field describing the content of statistical data as you have mentioned. Either a manual entry or as default text the settings of eg. OpenTimeTable (dateperiod, dayselection, timeslot, ...) could be written in this field.

Best regards  
Markus

> For example, if you have the mean departure time of March and the mean  
> departure time of calendar week 11 in the same file, an application that  
> is reading the file must be able to make a distinction. Whether this is  
> based on an somewhat arbitrary identifier ("DatasetMarkus418b" or  
> "MonthlyStatisticsMarch2005") or an explicit listing of all parameters  
> describing the dataset and its origin is - in my opinion - of minor  
> concern.

---