Subject: Delay Causes Representation in RailML
Posted by matteo.anelli on Mon, 08 Oct 2012 14:39:14 GMT
View Forum Message <> Reply to Message

Hi,
I'm Matteo Anelli and I'm working on the On-Time European Project (http://www.ontime-project.eu/)
on behalf of NTT Data.

The project aims to use RailML to model the base data structures of the project, to allow easier
sharing of information and data by distributed systems.

Working on the data modeling of the delay, I noticed that the entity does not include any field to model
the reason of delay.

Since the project want to integrate the FICHE UIC 450-2 delay causes (a two-digit number) I'd like to
see some extended attributes on the Delay Entity. I think a generic attributes field should be sufficient,
if not a string field for "causes". It's that possible?

Thanks

--
----== posted via PHP Headliner ==----

Subject: Re: Delay Causes Representation in RailML
Posted by Susanne Wunsch railML on Wed, 10 Oct 2012 11:42:47 GMT
View Forum Message <> Reply to Message

Hi Matteo and others interested,

Thanks for your introduction and for sharing your knowledge with the railML community.

matteo.anelli@nttdata.com (Matteo Anelli) writes:
> Since the project want to integrate the FICHE UIC 450-2 delay causes
> (a two-digit number) I'd like to see some extended attributes on the
> Delay Entity.

I don't really like the idea of the two-digit code for several reasons.

* The code in the UIC document may change (even with a very low

likelihood).

* In order to fill out or understand the attribute one has to purchase
the UIC document for currently 178 EUR.

> I think a generic attributes field should be sufficient,

That only helps out for the moment where the information should be transferred between two systems that explained their special additions each other.

> if not a string field for "causes".

That would enable the semantically fixed right position in the railML tree for data exchange. But "open string fields" entail the risk that two systems define the same meaning with different words or only different syntax.

> It's that possible?

The string field and the generic extension would be possible as quick fix. Nevertheless I would prefer an approach to define the delay causes as extensible enumeration lists.

I would presume there could be multiple causes to be defined for a single delay. Am I right?

Could you provide a suggestion very soon? (We currently feature freeze the new 2.2 release)

Thanks again and kind regards...
Susanne

--
Susanne Wunsch
Schema Coordinator: railML.common

Subject: Re: Delay Causes Representation in RailML
Posted by on Thu, 11 Oct 2012 20:21:28 GMT
View Forum Message <> Reply to Message

> Nevertheless I would prefer an approach to define the delay causes
> as extensible enumeration lists.

I agree.

> I would presume there could be multiple causes to be defined for a

> single delay. Am I right?

Not basically. If you allow multiple reasons you should also create a possibility to note the basic (earliest) reason - the other's being caused by that one. Possibly a time-stamp per reason could be useful - time when the certain reason was added. The basic one is the one with the earliest time-stamp.

I would suggest to allow multiple reasons in a kind they could easily be extended later (e. g. enumerable element).

Best regards,
Dirk.

Subject: Re: Delay Causes Representation in RailML
Posted by Susanne Wunsch railML on Mon, 15 Oct 2012 14:47:32 GMT
View Forum Message <> Reply to Message

Hi Dirk, Matteo and others,

(I just filed a Trac-ticket for this topic [1]).

Dirk Bräuer <dirk.braeuer@irfp.de> writes:
>> I would presume there could be multiple causes to be defined for a
>> single delay. Am I right?
>
> Not basically. If you allow multiple reasons you should also create a
> possibility to note the basic (earliest) reason - the other's being
> caused by that one. Possibly a time-stamp per reason could be useful
> - time when the certain reason was added. The basic one is the one
> with the earliest time-stamp.
>
> I would suggest to allow multiple reasons in a kind they could easily
> be extended later (e. g. enumerable element).

It's a bit hard to find an lightweight way to define these interactions.

The delay cause could be defined very deep inside the railML tree:

timetable/trainParts/trainPart/ocpsTT/ocpTT/statistics/stati stic/...

...mean/@arrivalDelay ...mean/@departureDelay
...median/@arrivalDelay ...median/@departureDelay
...standardDeviation/@arrivalDelay ...standardDeviation/@departureDelay

I just realized that there is no position to define the "real/actual/current" delays. Whereas the "real/actual/current" arrival and

departure times (and days) are categorized by 'scope="actual"'.

I would not recommend using the 'mean' element therefore, no matter that the mean value of one value is the same as the value itself. ;-)

There is a similar situation for the "real/actual/current" stop time

timetable/trainParts/trainPart/ocpsTT/ocpTT/...
...stopDescription/stopTimes/@minimalTime

that is a clear planning value, not fitting the statistical point of view.

How about extending the 'statistic' element by ...

...current/@arrivalDelay
...current/@departureDelay
...current/@stopTime

for the above discovered issues.

Coming back to the original purpose - proposing a position and structure for several more or less interdependent delay causes.

* new element 'delayCauses' as a container and sub-element to the newly proposed 'current' element

* one or more sub-elements 'delayCause' with basic railML attributes (id, description...) and a 'date' and 'time' attribute for the occurence history

* choice of the following sub-elements:
'infrastructureManagerResponsible', 'railwayUndertakingResponsible', 'otherResponsible' (following the code structure of UIC 450-2)

* depending on the '*Responsible' element there is a list of possible sub-elements with according cause attributes (also following the code structure of UIC 450-2 in an expandable matter)

The resulting structure may look like the following:

```
<current arrivalDelay="PT2M">
  <delayCauses>
    <delayCause id="dc1" time="17:12:00Z" date="2012-10-14"
      description="Something really special happened">
      <otherResponsible>
        <external subject="outsideInfluence"/>
      </otherResponsible>
```

```
    </delayCause>
  </delayCauses>
</current>
```

Further causes would be:

```
  <infrastructureManagerResponsible>
    <management
      subject="timetableCompilation"/
               "formationOfTrain"/
               "mistakesInOperationalProcedures"/
               "wrongApplicationOfPriorityRules"/
               "staff"/
               "other:foo"/>
    <infrastructureInstallations
      subject="signalling"/
              "signallingAtLevelCrossings"/
              "telecommunication"/
              "powerSupplyEquipment"/
              "track"/
              "structures"/
              "staff"/
              "other:foo"/>
    <civilEngineering
      subject="plannedConstructionWork"/
              "irregularitiesInExecutionOfConstructionWork"/
              "speedRestrictionDueToDefectiveTrack"/
              "other:foo"/>
    <otherInfrastructureManager
      subject="next"/
              "previous"/>
  </infrastructureManagerResponsible>
  <railwayUndertakingResponsible>
    <commercial
      subject="exceedingTheStopTime"/
              "requestOfTheRailwayUndertaking"/
              "loadingOperations"/
              "loadingIrregularities"/
              "commercialPreparationOfTrain"/
              "staff"/
              "other:foo"/>
    <rollingstock
      subject="rostering"/
              "formationOfTrainsByRailwayUndertaking"/
              "problemsAffectingPassengerCoaches"/
              "problemsAffectingFreightWagons"/
              "problemsAffectingEngines"/
              "staff"/
```

```
            "other:foo"/>
    <otherRailwayUndertaking
      subject="next"/
            "previous"/>
  </railwayUndertakingResponsible>
  <otherResponsible>
    <external
      subject="strike"/
            "administrativeFormalities"/
            "outsideInfluence"/
            "weatherAndNaturalCauses"/
            "nextNetwork"/
            "other:foo"/>
    <secondary
      subject="emergencySituation"/
            "trackOccupation"/
            "turnAround"/>
            "connection"/>
            "furtherInvestigationNeeded"/
            "other:foo"/>
  </otherResponsible>
```

Sorry for this longish posting.

Kind regards...
Susanne

[1] https://trac.assembla.com/railML/ticket/170

--
Susanne Wunsch
Schema Coordinator: railML.common

Subject: Re: Delay Causes Representation in RailML
Posted by on Wed, 17 Oct 2012 18:35:01 GMT
View Forum Message <> Reply to Message

Dear Susanne and all others,

> I just realized that there is no position to define the
> "real/actual/current" delays. Whereas the "real/actual/current" arrival
> and departure times (and days) are categorized by 'scope="actual"'.
>
> I would not recommend using the 'mean' element therefore, no matter
> that
> the mean value of one value is the same as the value itself. ;-)

I agree.

In general, concerning 'scope="actual"':

I have some problem with these 'actual' (running) information in a schema called 'timetable'. I think the term 'timetable' implements that it is about planning only, not about 'actual' (running) information.

I am not sure whether it is possible to describe the aspects of 'actual' (running) information in a proper way with current RailML. We have some information about that but they are more dead bodies from RailML1...

But, this should not mean that such information should not be possible in RailML. I just feel a little bit uncomfortable if we now extend the 'actual' information in such a considerable way before discussing the general matter. As Andreas Tanner wrote: There would be necessary more to describe 'actual' information in the current timetable schema. Especially, there would be a need to describe the actual day. At least, this means to document that <trains>/<trainParts> which a 'scope="actual"' do need an <operatingDay> for only one day (bitMask with only one '1').

> <current arrivalDelay="PT2M">
> <delayCauses>
> <delayCause id="dc1" time="17:12:00Z" date="2012-10-14"

Well, I see that you want to code the actual day in a <delayCause>? Does this mean that you want to add this structure as a sub-structure of the 'planned' train?

So, <delayCause>.date has to be one from the <operatingDay> of the higher structure?

Is <delayCause>.time redundant to arrivalTime + arrivalDelay, isn't it? (If so, please do not tell Andreas!)

Can I describe a delay (time) without knowing the cause?

--
> description="Something really special happened">
> <otherResponsible>
> <external subject="outsideInfluence"/>

I do not see the reason for the element name "external" but anyway, I confess I do not understand all of it. I do not need to. I trust you! ;-)

--
> subject="timetableCompilation"/

The "subject" attribute of <management>, <infrastructureInstallations>, <civilEngineering> a. s. o. is just what I meant with

>> I would suggest to allow multiple reasons in a kind they could easily
>> be extended later (e. g. enumerable element).

So: Very good. I would not see a need to group them at <management>, <infrastructureInstallations>, <civilEngineering> a. s. o. From my side, it would be enough just to enumerate them all in one attribute. I think currently and in near, middle future nobody will know the reasons in such a detailed way... But of course it is theoretically ok!

Best regards,
Dirk.

Subject: Re: Delay Causes Representation in RailML
Posted by Susanne Wunsch railML on Thu, 18 Oct 2012 07:51:27 GMT
View Forum Message <> Reply to Message

Hi Dirk, Matteo and others,

Dirk Bräuer <dirk.braeuer@irfp.de> writes:
>> I just realized that there is no position to define the
>> "real/actual/current" delays. Whereas the "real/actual/current"
>> arrival and departure times (and days) are categorized by
>> scope="actual"'.

> In general, concerning 'scope="actual"':
>
> I have some problem with these 'actual' (running) information in a
> schema called 'timetable'. I think the term 'timetable' implements
> that it is about planning only, not about 'actual' (running)
> information.

I would expect that the whole railML file either covers planning data or "actual" (running) data. AFAIK the SBB saves all "actual" timetable data of a day in a single railML file.

> But, this should not mean that such information should not be possible
> in RailML. I just feel a little bit uncomfortable if we now extend the
> actual' information in such a considerable way before discussing the
> general matter. As Andreas Tanner wrote: There would be necessary more
> to describe 'actual' information in the current timetable
> schema. Especially, there would be a need to describe the actual
> day. At least, this means to document that <trains>/<trainParts>
> which a 'scope="actual"' do need an <operatingDay> for only one day
> (bitMask with only one '1').

+1

I feel uncomfortable with implementing the actual delays in another way
than the actual times.

>> <current arrivalDelay="PT2M">
>>   <delayCauses>
>>   <delayCause id="dc1" time="17:12:00Z" date="2012-10-14"
>
> Well, I see that you want to code the actual day in a <delayCause>?
> Does this mean that you want to add this structure as a sub-structure
> of the 'planned' train?

No, I think that this sub-structure can semantically only be used with
"actual" timetable data.

  <train processStatus="actual" ...>

> So, <delayCause>.date has to be one from the <operatingDay> of the
> higher structure?

No, the delay cause does not in any event rely on the train run itself.
Though the delay cause may be dated by some external event that happened
sometimes earlier than the actual run time. In most cases it happened on
the same day as the train runs.

> Is <delayCause>.time redundant to arrivalTime + arrivalDelay, isn't
> it? (If so, please do not tell Andreas!)

No, it depends on the delay cause event, not actually on the actual
train run time or position.

But another aspect would be redundancy of the delay causes for
subsequent stops. Maybe they shouldn't be that deep in the tree but
defined separated and referred to.

> Can I describe a delay (time) without knowing the cause?

Yes. The delay cause characteristics should be optional and extensible.

>>   description="Something really special happened">
>>   <otherResponsible>
>>   <external subject="outsideInfluence"/>
>
> I do not see the reason for the element name "external" but anyway, I
> confess I do not understand all of it. I do not need to. I trust you!
> ;-)

Thanks. ;-)

Some ideas from the UIC 450-2 for the above event:

- Activation of the emergency alarm by or for a third party
- Ejection of passengers from the train
- Police intervention in the train
- Intervention of emergency or medical services
- Intervention by fire service
- Collision with road vehicles
- Bomb alarm, suspect luggage
- Escaped animals
- Death or birth in the train
- Arson
- Malicious acts
- Demonstrations
- Vandalism
- Embankment fire

> So: Very good. I would not see a need to group them at <management>,
> <infrastructureInstallations>, <civilEngineering> a. s. o. From my
> side, it would be enough just to enumerate them all in one
> attribute. I think currently and in near, middle future nobody will
> know the reasons in such a detailed way... But of course it is
> theoretically ok!

That didn't come out of my head, it is summarized from the UIC 450-2 leaflet. There are special codes for these categories that some software should re-calculate if needed.

I didn't want to go the easiest way with introducing an attribute "uicDelayCauseCode" for two digits. ;-)

Kind regards...
Susanne

Subject: Re: Delay Causes Representation in RailML
Posted by on Thu, 18 Oct 2012 09:27:50 GMT
View Forum Message <> Reply to Message

Dear Susanne,

ok, I understood that

>>> <delayCause id="dc1" time="17:12:00Z" date="2012-10-14"

is not the date of the delay but the date of the cause. So it is not redundant to "planned arr/dep" + "delay arr/dep" and I do not need to fill it.

Ok, sorry for the misunderstanding - subject is closed from my side.

>> At least, this means to document that <trains>/<trainParts>
>> which a 'scope="actual"' do need an <operatingDay> for only one day
>> (bitMask with only one '1').

> I would expect that the whole railML file either covers planning data or
> "actual" (running) data. AFAIK the SBB saves all "actual" timetable data
> of a day in a single railML file.

This is very important. I think we should fix it either in a general attribute or at least in Wiki.

Best regards,
Dirk.

Subject: Re: Delay Causes Representation in RailML
Posted by Andreas Tanner on Fri, 19 Oct 2012 09:18:18 GMT
View Forum Message <> Reply to Message

Am 18.10.2012 09:51, schrieb Susanne Wunsch:
> Hi Dirk, Matteo and others,
>
> Dirk Bräuer<dirk.braeuer@irfp.de> writes:
>>> I just realized that there is no position to define the
>>> "real/actual/current" delays. Whereas the "real/actual/current"
>>> arrival and departure times (and days) are categorized by
>>> scope="actual"'.
>
>> In general, concerning 'scope="actual"':
>>
>> I have some problem with these 'actual' (running) information in a
>> schema called 'timetable'. I think the term 'timetable' implements
>> that it is about planning only, not about 'actual' (running)
>> information.
>
> I would expect that the whole railML file either covers planning data or
> "actual" (running) data. AFAIK the SBB saves all "actual" timetable data
> of a day in a single railML file.
>

Dear all,

just my 2ct:

I don't know what exact use case the original poster has in mind. I can imagine two:
- either, like the mentioned SBB use case, the information for a complete day is saved in one file, ex post, e.g. for statistics
- or, more or less real-time, fragmentary information is transmitted.

For the first purpose, I think that the timetable schema could be used (even if Dirk is right that it is not its intention). However, I would not use the OperatingPeriod construct to describe every single day. Rather I would allow to use the original operating periods from the planning data (e.g., "daily") and provide an additional attribute "runDay" that identifies the instance the delay refers to. This way, one could just "enrich" a timetable with actual run information.

At least for the second use case, I vote for reduced redundancy ;-) :

Here, delay information should go into its own hierarchy. A delay information should
- refer to a trainPart from the planning data
- refer to an OCPTT of the train run
- carry the ocptt-specific delay information.
- optionally, refer to a delay cause.

Once more (see the discussion on connections), this would imply that the OCPTT need their own IDs.

<delayInfos>
<delayInfo trainPartRef = ... runDay=...>
<delay ocpttref=zzz current=...>
 <delayCauses><delayCauseRef ref=.../><delayCauseRef ref=.../></delay>
<delay ... />
</delayInfo>
</delayInfos>

<delayCauses>
<delayCause id=... />
....

About terminology: I would prefer <delayEvent> to <delayCause>.

--Andreas.

Subject: Re: Delay Causes Representation in RailML
Posted by Andreas Tanner on Fri, 19 Oct 2012 09:43:28 GMT

> Rather I would allow to use the original operating periods from the
> planning data (e.g., "daily") and provide an additional attribute
> "runDay" that identifies the instance the delay refers to.

The attribute could be anchored at the <times> element:

<ocpTT ocpRef=...>
<times scope="actual" runDay="01.01.2012" ... />
<times scope="actual" runDay="02.01.2012" ... />

Then, "runDay" should only be allowed for times with scope actual to avoid that it is abused as a replacement for OperatingPeriods.

This way, one
> could just "enrich" a timetable with actual run information.
>
... even with information for multiple days within one file, with clear mapping to the planning data.

--A.

Subject: Re: Delay Causes Representation in RailML
Posted by Joachim Rubröder railML on Wed, 07 Nov 2012 15:38:16 GMT

Hello everybody,

Matteo Anelli wrote:
> Since the project want to integrate the FICHE UIC 450-2 delay causes (a
> two-digit number) I'd like to
> see some extended attributes on the Delay Entity. I think a generic
> attributes field should be sufficient,
> if not a string field for "causes". It's that possible?

Let me try to summarize the current discussion related to the delays.

1. There is currently no entity for a 'delay' event in the timetable but several '...Delay' attributes within the statistic element
2. A definit list of delay causes would be preferable, but is not obvious to define
3. The timetable schema was not intended for, but could be used to describe real ex post data for statistical purposes, as it is done by SBB with OTT
4. There is not yet a consensus about the proper way of describing delay events

I would therefore suggest to provide an additional "any"-attribute for the elements with delay information (mean, median, standardDeviation) as immediate quickfix measure for the version 2.2 and keep the discussion running to get an idea of the appropriate way of describing current data and delays in a later version 2.3 or 3.0 of railML.

Are there any objections?

Kind regards,
Joachim

Joachim RubrÃ¶der
Schema Coordinator: railML.timetable

--
----== posted via PHP Headliner ==----

Subject: Re: Delay Causes Representation in RailML
Posted by Susanne Wunsch railML on Tue, 27 Nov 2012 11:40:19 GMT
View Forum Message <> Reply to Message

Hello Joachim, Matteo and others,

At the railML-conference in Zurich we agreed on discussing the aspect of "actual/real-time" data in a work group after publishing the 2.2 release.

Nevertheless the current railML schemas provide some possibilities for "actual" data - e.g. the ocpTT/times/@scope attribute.

```
<ocpTT ...>
  <times scope="scheduled" arrival="10:15:00" departure="10:25:00"/>
  <times scope="actual" arrival="10:20:17" departure="10:24:13"/>
</ocpTT>
```

That means that the train part arrived ca. five minutes too late and departed ca. one minute too early.

coord@timetable.railml.org (Joachim Rubröder) writes:
> 1. There is currently no entity for a 'delay' event in the timetable but
> several '...Delay' attributes within the statistic element

For simple arrival and departure delays the 'statistics' element is not needed, as shown above.

> 2. A definit list of delay causes would be preferable, but is not
> obvious to define

We agreed on an free string field, that may be pre-defined by a separate XML file from the railML domain and/or pre-defined by a customer specific XML file. The UIC-Leaflet would be a good starting point.

> I would therefore suggest to provide an additional "any"-attribute for the
> elements with delay information (mean, median, standardDeviation) as
> immediate quickfix measure for the version 2.2

I would welcome an 'anyAttribute' in the 'times' element. As shown above, the 'statistics' elements are not needed for this kind of 'actual' data.

As discussed at the railML-conference in Zurich there are strong arguments against the exchange of 'statistics'-results because of the lack of transferring its algorithms and covered data set.

> and keep the discussion
> running to get an idea of the appropriate way of describing current data
> and delays in a later version 2.3 or 3.0 of railML.

http://trac.assembla.com/railML/ticket/170

Kind regards...
Susanne

--
Susanne Wunsch
Schema Coordinator: railML.common