

TeamRoom Plus

IBM TeamRo

This OPEN document will not be filed. It is being kept active.

Meeting about Meetings\OGF

Project	DFDL 1.0
Meeting Date	18-Oct-11 (Tues)
Meeting Time	15:00 - 16:00

Created by Steve Hanson on 09-Mar-11

Last Modified by Steve Hanson on 18-Oct-11

OGF DFDL Working Group Call , 18 October 2011

Agenda

Prepare for your meeting by describing the objectives (both immediate and long-term, if appropriate) of the meeting; and describe key details.

1. Action 136

This was closed recently, with the following conclusion:

"Favoured approach is to extend the concept of defaulting on parsing to include empty members of a sparse array "

But a better solution has since been identified, which involves the extension of nillable to complex types, but in a constrained way that only allows a nil literal value of %ES;. To maintain position in a sparse array, make the element nillable.

2. Further spec issues from IBM

As implementation has progressed, a number of issues have been spotted in the DFDL 1.0 spec by the IBM implementation team. Remaining items for resolution:

6.3. When a property can be either a DFDL String Literal or a DFDL Expression, the rules for recognizing the latter are:

- Must start with a '{' in the first position and end with '}' in the last position.
- '{' in any other position is treated as a literal
- '}' in any position other than the last position is treated as a literal.
- '{{' as the first characters are treated as the literal '{' and not a DFDL expression.

Question - should the following give a schema definition error for an invalid expression , or be treated as a string literal? The 1st rule says it can't be an expression.

Example: " {xxxxxxx"

However, this clearly would give a schema definition error for an invalid expression if the property could *only* be an expression. That seems inconsistent.

13.11. Should DFDL allow a date/time value to be silent about timezone?

Currently the dfdl:calendarTimeZone property is used to supply a TZ when there is none in the data.

But this means we are not compatible with XML Schema 1. where "no TZ" is an allowable state.

XML Schema 1.0 validation validates a date/time value against facets according to:

<http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/#dateTime-order>

In summary:

-> value with TZ & facet with TZ is a direct comparison
-> value with no TZ & facet with no TZ is a direct comparison
-> value with TZ & facet with no TZ uses a TZ range -14/+14
-> value with no TZ & facet with TZ uses a TZ range -14/+14
Should dfdl:calendarTimezone be enhanced to allow empty string meaning "no TZ" ?

13.11. Sandy Gao from XML Schema WG has noted that when unparsing a date/time infoSet item, DFDL applies the dfdl:calendarTimezone property. To match XML Schema, this would be applied after validation has taken place, if enabled. But the subsequent application of the property then changes the value in the data. By analogy, DFDL can apply rounding rules when unparsing, which can also change the value in the data, eg, '1.9999' gets rounded to '2.0' which could then fail validation if re-parsed. Is this a problem or just the way it works? Should this property be parsing only?

13.11. The DFDL property combination of dfdl:calendarTimezone and dfdl:calendarObserveDST is insufficient as it does not cater for different start dates for DST. To achieve this we would have to relax dfdl:calendarTimezone rules and allow it to take other formats, such as:

a) Olson format (eg, "**Europe/Paris**")

b) Unicode short time zone IDs as per ([Unicode Locale Data Markup Language](http://unicode.org/reports/tr35/)) spec (eg, "**frpar**")

The list of TZs for b) is here: <http://lawiki.org/extensions/core/common/bcp47/timezone.xml>

Note there a special value for 'unknown' - "**unk**" or "**Etc/Unknown**" - prefer to empty string?

13.11.1. DFDL only allows 'y' as year symbol in dfdl:calendarPattern. This only allows positive values, any negative value is ignored unless you also specify the 'G' (era) symbol, and there is no year 0 (which means that negative astronomical dates are one year out). DFDL could support the ICU 'u' extended year symbol which allows year 0 (means BC-1), corresponds to astronomical years and matches XML Schema 1.0 after errata applied for year 0.

3. AOB

Steve is away next Tuesday

Minutes

Meeting Minutes

Reflect on your meeting as you record all topics and issues discussed, and any tabled conversations. What went well, or what would you do differently next time? Document those so others can take advantage of your learning.

Attendees

Steve Hanson (IBM) (co-chair)

Mike Beckerle (co-chair)

Steph Fetzler (IBM)

Tim Kimber (IBM)

Apologies

Suman Kalia (IBM)

Minutes

1. Action 136

This was closed recently, with the following conclusion:

"Favoured approach is to extend the concept of defaulting on parsing to include empty members of a sparse array "

But a better solution has since been identified, which involves the extension of nillable to complex types, but in a constrained way that only allows a nil literal value of %ES;. To maintain position in a sparse array, make the element nillable.

Agreed that action 136 conclusion should be revised as above.

Need to work out the rules for nillable complex elements.
This will be done as part of action 140.

2. Further spec issues from IBM

As implementation has progressed, a number of issues have been spotted in the DFDL 1.0 spec by the IBM implementation team. Remaining items for resolution:

6.3. When a property can be either a DFDL String Literal or a DFDL Expression, the rules for recognizing the latter are:

- Must start with a '{' in the first position and end with '}' in the last position.
- '{' in any other position is treated as a literal
- '}' in any position other than the last position is treated as a literal.
- '{{' as the first characters are treated as the literal '{' and not a DFDL expression.

Question - should the following give a schema definition error for an invalid expression, or be treated as a string literal? The 1st rule says it can't be an expression.

Example: "{xxxxxxx}"

However, this clearly would give a schema definition error for an invalid expression if the property could *only* be an expression. That seems inconsistent.

Agreed - enforce use of {{ for all properties that can be either DFDL String Literal or expression.
Errata taken.

13.11. Should DFDL allow a date/time value to be silent about timezone?

Currently the dfdl:calendarTimeZone property is used to supply a TZ when there is none in the data. But this means we are not compatible with XML Schema 1. where "no TZ" is an allowable state.

XML Schema 1.0 validation validates a date/time value against facets according to:

<http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/#dateTime-order>

In summary:

- > value with TZ & facet with TZ is a direct comparison
- > value with no TZ & facet with no TZ is a direct comparison
- > value with TZ & facet with no TZ uses a TZ range -14/+14
- > value with no TZ & facet with TZ uses a TZ range -14/+14

Should dfdl:calendarTimezone be enhanced to allow empty string meaning "no TZ" ?

Agreed - allow empty string or another value depending on outcome of **action 153** below.

13.11. Sandy Gao from XML Schema WG has noted that when unparsing a date/time infoset item, DFDL applies the dfdl:calendarTimeZone property. To match XML Schema, this would be applied after validation has taken place, if enabled. But the subsequent application of the property then changes the value in the data. By analogy, DFDL can apply rounding rules when unparsing, which can also change the value in the data, eg, '1.9999' gets rounded to '2.0' which could then fail validation if re-parsed.

Is this a problem or just the way it works? Should this property be parsing only?

Agreed that this should be parsing only, on the grounds that the infoset should be providing either all the content for a value (so not relying on bits being added) or nothing in which case a default is used if required. Errata taken.

13.11. The DFDL property combination of dfdl:calendarTimezone and dfdl:calendarObserveDST is insufficient as it does not cater for different start dates for DST. To achieve this we would have to relax dfdl:calendarTimeZone rules and allow it to take other formats, such as:

a) Olson format (eg, "**Europe/Paris**")

b) Unicode short time zone IDs as per ([Unicode Locale Data Markup Language](http://unicode.org/reports/tr35/#Unicode_Locale_Data_Usage)) spec (eg, "**frpar**")

The list of TZs for b) is here: <http://lawiki.org/extensions/core/common/bcp47/timezone.xml>

Note there a special value for 'unknown' - "**unk**" or "**Etc/Unknown**" - prefer to empty string?

Agreed in principle, raised **action 153** to decide on the extended format to use.

Noted that calendarObserveDST would still be honoured for UTC values.

13.11.1. DFDL only allows 'y' as year symbol in dfdl:calendarPattern. This only allows positive values, any negative value is ignored unless you also specify the 'G' (era) symbol, and there is no year 0 (which means that negative astronomical dates are one year out). DFDL could support the ICU 'u' extended year symbol which allows year 0 (means 1BC), corresponds to astronomical years and matches XML Schema 1.0 after errata applied for year 0.

Agree to add 'u' symbol. There are other unsupported symbols but they are obscure and will not be added unless a need is shown. Errata taken.

3. AOB

Steve is away next Tuesday so no call next week.

Meeting closed

15:50 UK

Next call

Tues 1st Nov 2011

Create Action Items

Record the to-do's and individuals assigned by entering the appropriate information in the form below . Press the "Create Action Items" button to create specific to do's that can be tracked in the assignee's Work for Me views. " All Action Items will be tracked in the Action Items and Other Meeting Documents tab .

Action Items and Other Meeting Documents

View: ResultDocs

Next action: 154

Actions raised at this meeting

No	Action
153	Decide which TZ schema to adopt to extend dfdl :calendarTimeZone (Steve) 18/10: Need to keep as close to ICU as possible as an aid to implementers.

Current Actions :

Action

No	
123	<p>DFDL tutorial (Steve)</p> <p>13/10: Draft of first 3 chapters has been written and will be distributed to WG</p> <p>10/11: Posted to grid forge here (http://forge.gridforum.org/sf/go/doc16106?nav=1), work continuing at IBM to define a standard example-based chapter framework and to author additional chapters. Contributors welcome!</p> <p>17/11: Steve, Stephanie and Alan had a meeting to discuss the best structure for the tutorial and decide which examples to use throughout. The meeting raised more questions. Further discussions will be held.</p> <p>24:11: The list of topics to be covered in the remaining lessons has been produced and a lesson template. Alan will write lesson 4</p> <p>01/12: Alan has started lesson 4 which covers fixed and variable fields and arrays .</p> <p>08/12: Alan has almost completed lesson 4. Will send out for review.</p> <p>15/12: First draft of lesson 4 is available for review. Alan to send to Bob and Joe.</p> <p>22/12: Alan has distributed drafts for tutorials on Basic Structure and Optional/Repeating elements. Please review</p> <p>12/01: Alan distributed a tutorial for choices and updated the others. Alan and Steve reviewed them and updated versions will be sent soon. Should start on the 'representation' tutorials soon.</p> <p>19/01: The tutorials for basic structure, optional/arrays and choices have be updated. Please review. The tutorial for text elements should be available soon.</p> <p>26/01: No comments received about 3 tuorials distributed last week. Alan is still working on Text representation.</p> <p>02/02: Steve has sent comments on three tutorials. Alan to send updated versions by the end of the week. Alan has also distributed the first part of the tutorial on text representation and would like feedback.</p> <p>09/02: Steve had reviewed tutorials 3,4,5 and updated versions have been distributed. Joe reviewed lesson on text elements.</p> <p>Main points. Using 'represented as text' is confusing. Examples are too cluttered. Suggest simple targeted examples but still build up to final complete schema</p> <p>23/02: New versions distributed and Steve has commented.</p> <p>02/03: Alan has published the final versions of tutorials 4,5,6 and is working on text representations. There was some discussion about the detail that needs to be covered. Should limit it to 'common usage' and refer to the spec for details of edge cases.</p> <p>09/03: Alan distributed an update to the text tutorial. Please review.</p> <p>30/03: Steve has spent half a day tidying up lessons 1 to 6 and has uploaded them as pdfs to gridforge. They are now more coherent, and many inconsistencies and errors fixed. Ownership of draft lessons (text properties, binary properties, advanced features) has been passed to Steve. Also need to make a schema available for the examples.</p> <p>13/04: Steve is working on the text properties tutorial.</p> <p>04/05: No progress</p> <p>18/05: No progress</p> <p>01/06: No progress</p> <p>08/06: No progress</p> <p>15/06: This is on hold until Steve clears up spec issues and other workload . Steph has looked at the later lessons, and noted that they are more direct compared to the more wordy earlier lessons.</p> <p>28/06: On hold.</p> <p>...</p> <p>18/10: On hold</p>
132	<p>Publishing DFDL xsd (Suman)</p> <p>08/12: Agreed that it should be made available. Suman has started the approval process in IBM</p> <p>15/12: no progress</p> <p>22/12: no update</p> <p>12/01: Suman is getting approval from IBM to publish.</p> <p>19/01: Waiting to get IBM approval to publish</p> <p>26/01: no update</p> <p>02/02: No update</p> <p>09/02: no update</p>

23/02: no update
 02/03: Suman is working through the IBM process to permit publication. There was discussion about what licence the XSD would be published under and how that would effect use in products. Suman to investigate
 09/03: No update
 30/03: Suman has sent information to IBM legal. Reminded him about the license issue.
 13/04: No update.
 04/05: IBM has permission to release the DFDL model xsds to WG members only, Suman has a couple more changes to make and will send to Steve for review. License clarification needed.
 18/05: Awaiting response from IBM legal. Suman will send Steve the model xsd for review.
 01/06:
 08/06: Awaiting response from IBM legal. Suman will send Steve the model xsd for review.
 15/06: Steve has received the xsds (there are three of them) and will review.
 28/06: Not reviewed yet
 05/07: Need to fully understand what WG members are able to do with it. The real usefulness is in other implementers being able to use the xsds to validate DFDL xsds so the license needs to reflect this.
 12/07: No further progress
 19/07: Suman will talk to IBM legal and make it clear that the license must allow users to actively use the xsds.
 26/07: No update
 16/08: Steve has reviewed and sent comments to Suman. Suman has asked legal to clarify re-distribution rules, taking into account OGF licensing terms and patent filing.
 23/08: No update
 30/08: Steve has reviewed words from IBM legal. It looks like that once the xsds are published by OGF then they are free to be used under the normal OGF licensing terms. Steve needs to clear all this with IBM legal in Hursley, as the DFDL mission is owned by Hursley in UK.
 20/09: Steve has mailed IBM Hursley legal
 27/09: Steve to meet with IBM Hursley legal
 4/10: No further progress
 18/10: Steve has met with legal, can publish the xsds after tidy up. Will create OGF info document.

133

Make a set of default formats available (Suman)
 19/01: Suman expects some default formats to be ready by Feb 9th. Will need approval to publish
 26/01: Stephanie sent the defaults used by test cases to Suman
 02/02: no update
 09/02: no update
 23/02: no update
 02/03: Same as 132
 09/03: No update. Same license issue apply though.
 13/04: No update.
 04/05: IBM will make one default format available. Suman is working through the IBM process to permit publication.
 18/05: Awaiting response from IBM legal.
 01/06:
 08/06: Awaiting response from IBM legal. IBM also want to prove that the default format has the properties sensibly defined to plans to include in internal testing.
 15/06: No change.
 ...
 19/07: 19/07: Suman will talk to IBM legal and make it clear that the license must allow users to actively use the format xsd(s).
 26/07: No update
 16/08: Same issue as action 132
 23/08: No update
 30/08: As action 132
 20/09: Steve has mailed IBM Hursley legal
 27/09: Steve to meet with IBM Hursley legal
 4/10: No further progress

	<p>18/10: Steve has met with legal, can publish the xsds after tidy up. Will create OGF info document.</p>
140	<p>Spec issue: Parsing: 'missing' v 'empty', role of initiators, default values (All)</p> <p>01/06: See minutes.</p> <p>08/06: Still under discussion. Tim has sent Mike a selection of data formats to guide the discussion.</p> <p>15/06: Not discussed - an extra call has been scheduled to go through this.</p> <p>28/06: A series of extra calls are being held between Mike, Steve, Tim and Steph.</p> <p>05/07: Next extra call is Wed 6th July - Steve to send invite</p> <p>12/07: Two more calls held. Next call is Wed 13th July.</p> <p>19/07: More calls held, next call is Fri 22nd July.</p> <p>26/07: More calls held, good progress</p> <p>16/08: Steve will set up next call for when Tim has returned from holiday</p> <p>23/08: Two more calls scheduled for this week, remaining issues: separator suppression, empty strings, sparse arrays (see action 136)</p> <p>30/08: Call held earlier today. Still remaining - separator suppression (matrix); sparse arrays; empty strings; empty value delimiter policy. Steve to summarise where we have got so far before remaining items are tackled.</p> <p>20/09: Steve has summarised where we are with action 140, which Tim and Mike have reviewed. 2 hours call planned for Thursday.</p> <p>27/09: Calls held, progressing the separator suppression behaviour</p> <p>4/10: More calls held, progressing separator suppression, sparse arrays (see action 136) and emptyValueDelimiterPolicy behaviour</p> <p>18/10: All issues now in a single document, call held earlier today. Next call Thursday.</p>
145	<p>Provide a 'dispatch' way of discriminating a choice for better performance of the envelope/payload use case (Steve, Mike, Suman)</p> <p>12/7: See minutes. Need to choose a proposal and flesh out.</p> <p>19/07: Waiting for proposals</p> <p>26/07: Waiting for proposals</p> <p>16/08: Waiting for proposals. Suman added to action.</p> <p>...</p> <p>18/10: Steve to send a proposal</p>
146	<p>Model NRL's bit-oriented format where complex repeating element carries a 'last' indicator (Steve, Mike)</p> <p>19/07: NRL to send xsd and example of format.</p> <p>26/07: Steve/Tim put forward a proposal that used a discriminator that looked at the previous item's repeatBit in the array and failed if it was 0. This looks like it will work. Mike to add the correct DFDL annotations to take into account both repeatBit and presentBit in a single discriminator, and auto-set them using outputValueCalc.</p> <p>A more usable solution to the 'last indicator' problem is desirable though. Noted that this has a 'repeat..until' semantic.</p> <p>16/08: Adam/Ryan to evaluate Steve/Tim proposal as it is potentially fragile relying as it does on backtracking.</p> <p>23/08: Adam/Ryan had meeting with Boyd Fletcher and others, still evaluating to see whether change to spec is needed to handle this better.</p> <p>30/08: No update</p> <p>20/09: Looking like a spec change would be beneficial. Adam will provide details within the next month.</p> <p>27/09: Waiting to hear from NRL. Mike to find out how they are progressing.</p> <p>4/10: Mike will contact NRL to see how things are progressing.</p> <p>18/10: No update - Steve & Mike to chase</p>

Closed actions

No	Action

Deferred actions

--

No	Action
129	<p>Press release to publicise DFDL (Steve) Steve is pulling together a press release at IBM. Want to include as many contributors and interested parties as possible. NCSA are keen to be included. Also likely that US National Archive will want to be included. Mike has indicated OCO are too. 17/11: no progress ... 08/12: Still no response from IBM press office 15/12: no progress 09/03: No progress 30/03: Making this action deferred until IBM is in a position to say something more concrete about any implementation.</p>
131	<p>Transformation of DFDL properties to a canonical form (Joe) 08/12: Joe has produced a XSLT to transform a DFDL schema to a canonical element form. When tested it should be made available on the WG gidforge site. 15/12: Alan tested against test dfdl schema which worked correctly (after fixing some errors in the schema) 22/12: no update 12/01: Joe has some defects to fix before making available on gridforge. 19/01: There is a difficult problem to solve before Joe make the style sheet public 26/01: Working on problems 02/02: no progress 09/02: As it wasn't a simple as exoected this will be treated as a low priority action 23/02: Low prioity 09/03: Low priority 30/03: Deferring for now</p>
066	<p>Investigate format for defining test cases (All) 25/11: IBM to see if it is possible to publish its test case format. 04/12: no update ... 17/02: IBM is willing in principle to publish the test case format and some of the test cases. May need some time to build a 'compliance suite' 24/03: No progress 03/03: Discussions have been taking place on the subset of tests that will be provided. 10/03: work is progressing 17/03: work is progressing 31/03: work is progressing 14/04: And XML test case format has been defined and is being tested. 21/04. Schema for TDML defined. Need to define how this and the test cases will be made public 05/05: Work still progressing 12/05: Work still progressing 02/06: Work still progressing on technical and legal considerations ... 25/08: Will chase to allow Daffodil access to test cases. The WG should define how implementation confirm that they 'conform to DFDL v1' 01/09: IBM still progressing the legal aspect. Intends to publish 100 or so tests as soon as it can, ahead of a full compliance suite. 08/09: IBM still progressing 15/09: IBM still progressing, expect tests to be available within a few weeks 22/09: IBM still progressing, expect tests to be available within a few weeks 29/09: Test cases are being prepared. 06/10: Some test cases should be available next week. Steve would like to be able to show the test case information at OGF 30. 13/10: Still progressing 10/11: Legal issues cleared, IBM in process of collecting 100 example test cases, ideally ones that fit the 'extended conformance' of NCSA Daffodil 17/11: Work is progressing on verifying the test cases. It should be possible to distribute to the WG in 2 weeks.</p>

24/11: About half the test cases have been completed and are being reviewed internally.
01/12: Test cases should be available shortly
08/12: The test cases are in internal IBM review. Probably need a bit of reorganising before publication
Stephanie gave a brief overview of the format of the test cases.
15/12: Ruth joined the call to provide the latest status. The test cases have been updated and a draft read.me produced. Although not ready for public distribution Ruth will send them to Joe for feedback.
22/12: Test cases were sent to Joe for initial testing which found some problems in the Daffodil parser
12/01: All current tests use a default format which Daffodil doesn't currently support. Joe suggested that there should be test that defined the same function using different definition forms. Also suggested that default formats should be provided by the WG. This had always been the intention. Action 133 raised to track.
19/01: There is currently no resource available in IBM to make more tests available. IBM to discuss how/if it can make a 'minimal compliance test suite' available.
26/01: Action kicked off within IBM. There was a brief discussion about naming and organisation of test cases but no preferences were expressed
02/02: IBM will not have the resources to develop a full test suite in the near future. Steve suggested that we produce a list of required test cases so that anyone could supply them.
09/02: Steve had previously sent a list of areas to be tested. Please review.
23/02: Please review Steve's list of areas to be tested
02/03: Alan had reviewed Steve's list and we went through his comments. Agreed there is no need for separate tests for the infoset or for dfdl: property lists, unions etc but comment will be added that these should be exercised during property testing.
09/03: Alan updated the test document. Need more introduction and perhaps adopting the OGF template.
30/03. Ownership of test document passed to Steve. This action is merged with 112 and will cover all aspects of compliance suite.
13/04: IBM will not have time to create a compliance suite in the near future. Probably best to make this action deferred for now.

Work items:

No	Item	Owner	Target	Status
043	Track errata list for 1.0 of the spec. http://forge.gridforum.org/sf/go/doc16280?nav=1	Steve	N/A	Draft 0004 on grid forge.
044	Incorporate errata list into DFDL spec.	Steve	TBD	