

TeamRoom Plus

IBM TeamRoom

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

Meeting about Meetings\OGF

Project	DFDL 1.0
Meeting Date	27-Sep-11 (Tues)
Meeting Time	15:00 - 16:00

Created by Steve Hanson on 09-Mar-11

Last Modified by Steve Hanson on 27-Sep-11

OGF DFDL Working Group Call , 27 September 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. 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.

13.6. Property naming convention. The convention is that the letters in a property name are lower case, with 2nd and subsequent words starting with upper case, even if the letters form an acronym, eg, utf16Width. However:

- calendarObserveDST has DST all upper case
- IBM implementation spells textStandardNaNRep as textStandardNaNRep, which is the natural way to want to write NaN. Should spec change or not?

13.15. When nilKind is 'literalValue', there is a restriction that the element must have text representation. This was added in draft 38, but Steve can find no record of why it was added in the WG call minutes at that time. Why was the done? Is it too restrictive? If this was relaxed, what values could the DFDL String Literals take?

14.1. Complex types with empty sequences. Spec correctly says that XML Schema rules mean a DFDL annotation on such a sequence is ignored, but also says it is a schema definition error. Many schema processors will not even inform the consumer that there was an empty sequence in the first place. So should say that an empty complex type gives a schema definition error.

2. Validation of DFDL schemas

Section 2.2 says "All conforming DFDL processors must detect all schema definition errors..". That's a big task for an implementation.

- What if the implementation only implements a parser, should it validate unparsing only properties?
- What if the implementation does not implement optional features?
- Is this true even if the schema definition error is not encountered when parsing/unparsing the actual data?
- What is the recommendation for global elements/groups, where properties may be completed by element/group refs?
- What is the recommendation for simple types, where properties are combined with the using element?
- What if a benign but non-applicable property is explicitly supplied on an object?

3. AOB

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)
Suman Kalia (IBM)
Tim Kimber (IBM)
Mike Beckerle (Deloitte)

Apologies

1. 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.

13.6. Property naming convention. The convention is that the letters in a property name are lower case, with 2nd and subsequent words starting with upper case, even if the letters form an acronym, eg, utf16Width. However:

- calendarObserveDST has DST all upper case
- IBM implementation spells textStandardNaNRep as textStandardNaNRep, which is the natural way to want to write NaN.

Should spec change or not?

Keep calendarObserveDST as per spec, but change textStandardNaNRep to 'NaN'.

13.15. When nilKind is 'literalValue', there is a restriction that the element must have text representation. This was added in draft 38, but Steve can find no record of why it was added in the WG call minutes at that time.

Why was it done? Is it too restrictive? If this was relaxed, what values could the DFDL String Literals take?

Allow representation to be binary, but restrict the content to be only %r or %ES

14.1. Complex types with empty sequences. Spec correctly says that XML Schema rules mean a DFDL annotation on such a sequence is ignored, but also says it is a schema definition error. Many schema processors will not even inform the consumer that there was an empty sequence in the first place. So should say that an empty complex type gives a schema definition error.

Agreed

2. Validation of DFDL schemas

Section 2.2 says "All conforming DFDL processors must detect all schema definition errors..". etc.

The spec mandates that the objects encountered by the DFDL processor during a parse or unparse are checked for schema definition errors. But that's as far as it goes. Anything beyond that is up to the implementation, due to the nature of the DFDL scoping rules..

What if the implementation only implements a parser, should it validate unparsing only properties?

- Up to implementation.

What if the implementation does not implement optional features?

- Up to implementation.

What is the recommendation for global elements/groups, where properties may be completed by element/group refs?

- Up to implementation. A suggestion is only validate global elements if not referenced.

What is the recommendation for simple types, where properties are combined with the using element?

- Only need to validate them independently if used as the target of prefixLengthType property.

What if a benign but non-applicable property is explicitly supplied on an object that is encountered during a parse or unparse?

- Discussed whether this should be a schema definition error but decided that a warning allows for easier composition. Will add to spec.

Meeting closed

16:05 UK

Next call

Tues 4th Oct 15:00 UK

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: 153

Actions raised at this meeting

No	Action

Current Actions :

No	Action
123	DFDL tutorial (Steve) 13/10: Draft of first 3 chapters has been written and will be distributed to WG 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! 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. 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>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>27/09: 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> <p>23/02: no update</p> <p>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</p> <p>09/03: No update</p> <p>30/03: Suman has sent information to IBM legal. Reminded him about the license issue.</p> <p>13/04: No update.</p> <p>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.</p> <p>18/05: Awaiting response from IBM legal. Suman will send Steve the model xsd for review.</p> <p>01/06:</p>

	<p>08/06: Awaiting response from IBM legal. Suman will send Steve the model xsd for review.</p> <p>15/06: Steve has received the xsds (there are three of them) and will review.</p> <p>28/06: Not reviewed yet</p> <p>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.</p> <p>12/07: No further progress</p> <p>19/07: Suman will talk to IBM legal and make it clear that the license must allow users to actively use the xsds.</p> <p>26/07: No update</p> <p>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.</p> <p>23/08: No update</p> <p>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.</p> <p>20/09: Steve has mailed IBM Hursley legal</p> <p>27/09: Steve to meet with IBM Hursley legal</p>
133	<p>Make a set of default formats available (Suman)</p> <p>19/01: Suman expects some default formats to be ready by Feb 9th. Will need approval to publish</p> <p>26/01: Stephanie sent the defaults used by test cases to Suman</p> <p>02/02: no update</p> <p>09/02: no update</p> <p>23/02: no update</p> <p>02/03: Same as 132</p> <p>09/03: No update. Same license issue apply though.</p> <p>13/04: No update.</p> <p>04/05: IBM will make one default format available. Suman is working through the IBM process to permit publication.</p> <p>18/05: Awaiting response from IBM legal.</p> <p>01/06:</p> <p>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.</p> <p>15/06: No change.</p> <p>...</p> <p>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).</p> <p>26/07: No update</p> <p>16/08: Same issue as action 132</p> <p>23/08: No update</p> <p>30/08: As action 132</p> <p>20/09: Steve has mailed IBM Hursley legal</p> <p>27/09: Steve to meet with IBM Hursley legal</p>
136	<p>Arrays with missing elements (Steve)</p> <p>There is a problem when there are empty/missing array elements with an index greater than minOccurs. For example:</p> <p>xs:element name="array" minOccurs=0, maxOccurs=10 lengthKind='delimited'</p> <p>Datastream: ,,,value3,value4</p> <p>InfoSet will contain:</p> <p>array[1] = value3</p> <p>array[2] = value4</p> <p>This is because elements with an index greater than minOccurs are optional and so do not get defaulted.</p>

	<p>Unparsing this info set will produce:</p> <p>Datastream: value3,value4</p> <p>You could make the empty space (%ES;) the nil value which will work for simple elements but not for complex.</p> <p>Info set will contain:</p> <pre>array[1] = nil array[2] = nil array[3] = value3 array[4] = value4</pre> <p>23/02: Discussed options</p> <ol style="list-style-type: none"> 1. Changed definition of required for arrays to be 'required up to the last instance in the data stream of the array' 2. add index to the element info item <p>Steve to investigate if XDM uses an index.</p> <p>02/03: No progress</p> <p>09/03: No update</p> <p>30/03: Stephanie recognised the issue from IBM's WTX. Here, the solution was to provide an option so that the user explicitly chose whether the position in the array was significant, as it is not always and on output some users do not want defaults or blanks to appear. Not yet resolved.</p> <p>13/04: Steve has verified that XDM does not carry index information, but will check with IBM's W3C rep to see if that has ever been expressed as a requirement. Steve also said that the area of defaulting missing required elements on parsing (especially complex elements) is one that the IBM implementation team has raised some concerns about, so Steve and the team are looking at this area again. It is possible that the spec will change to clarify behaviour, and so this action should be used to cover this work.</p> <p>04/05: In progress.</p> <p>18/05: In progress, at minimum some clarification to the spec is needed</p> <p>01/06: Discussed to bring Mike up to speed. Solution will depend on other spec discussions.</p> <p>08/06: Still under discussion</p> <p>15/06: Will come back to this after action 140 resolved</p> <p>...</p> <p>27/09: See action 140.</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>
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>

	16/08: Waiting for proposals. Suman added to action. ... 27/09: Steve to send a proposal
146	Model NRL's bit-oriented format where complex repeating element carries a 'last' indicator (Steve, Mike) 19/07: NRL to send xsd and example of format. 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. A more usable solution to the 'last indicator' problem is desirable though. Noted that this has a 'repeat..until' semantic. 16/08: Adam/Ryan to evaluate Steve/Tim proposal as it is potentially fragile relying as it does on backtracking. 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. 30/08: No update 20/09: Looking like a spec change would be beneficial. Adam will provide details within the next month. 27/09: Waiting to hear from NRL. Mike to find out how they are progressing.
147	Clarify the rules when for padding /trimming and nil handling interact (Steve, Mike) 19/07: It would seem that for literalCharacter you would want to apply nil processing first, else the nil character can get trimmed away. But for literalValue, trimming first is desirable to allow a single nil value to apply to multiple different fixed length elements. There is an inconsistency here, but is that ok? 26/07: No further thought on this 16/08: No further thought on this. Needs sub-team to evaluate. 23/08: No progress, will be discussed when current agenda items are completed 30/08: No progress, will be discussed when current agenda items are completed ... 27/09: As above

Closed actions

No	Action
148	Clarify the rules around pattern -based lengths and scanability (Mike) 26/07: Mike to provide words 16/08: Mike and Steve have discussed via e-mail and will report back to WG when concluded. ... 20/09: Discussed at length. The conservative approach of only allowing patterns with text data is favoured. Steve to write up words for 12.3.5.1. 27/09: Closed. Approved Steve's wording.
152	Decide whether calendarTimeZone property is modelled as an enum or as a pattern . 20/09: Suman to send two simple types, one showing pattern and the other the enums. 27/09: Closed. Use a pattern, allow only UTC and not GMT.

Deferred actions

No	Action
129	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.
131	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
066	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. 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 whih 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 info set 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	Ongoing
044	Incorporate errata list into DFDL spec.	Steve	TBD	