

# 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	19-Jun-12 (Tues)
Meeting Time	15:00 - 16:00

Created by Steve Hanson on 09-Mar-11

Last Modified by Steve Hanson on 18-Jun-12

## OGF DFDL Working Group Call , 18/19 June 2012

### Agenda

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

#### 1. Daffodil Open Source Project

Status.

#### 2. Does DFDL need to support signed integer types with lengthUnits 'bits'?

Spec only allows unsigned integers to have lengthUnits 'bits', but it is possible in C structures to declare bit fields for signed integers. They behave like two's complement. It looks like the spec is being too restrictive in its types for 'bits'.

#### 3. Unsupported style of zoned decimal overpunching encountered

On HP NonStop Tandem architecture, the negative sign is incorporated in the last byte of the number but using the highest bit (ie, 8) of the nibble, instead of the second highest bit (ie, 4). So instead of ASCII characters x70 to x79 representing negative 0-9, characters xB0 to xB9 are used. Accordingly we should add a new enum to support this, eg, 'asciiTandemModified'. If we do this, then bearing in mind that the positive bit pattern for the high order nibble is 0011 (ie, x3), DFDL will support the following overpunching:

- set the 8 bit (1011) => xB0 - xB9 (Tandem)
- set the 4 bit (0111) => x70 - x79 (standard)
- unset the 1 bit (0010) => x20 - x29 (CA Realia)

#### 4. Should parser check for parent 's terminating delimiters after parsing each child ?

See email thread. If my record format was some required fixed length fields followed by some optional fixed length fields with an indicator for end of record, I could not model the record as lengthKind 'delimited' (unless I used asserts), I would need to use lengthKind 'pattern'. That is as designed, but perhaps worth revisiting whether in-scope parent delimiters should be checked before parsing each child element?

#### 5. Agree on signature for new DFDL function to construct a DFDL string literal

See email. Should we also change DFDL function names to be hyphenated to match XPath's?

#### 6. ICU exponent symbol behaviour .

Errata 2.26.says "Disallow the use of empty string for property textStandardExponentCharacter, and state property must be set if the pattern contains 'E' (schema definition error otherwise)". An intentional side-effect of this rule is that when the pattern does \*not\* contain the exponent symbol, a DFDL processor can use an ICU API to switch off the use of exponent symbol, so that an exponent in the data will give an error when parsing. This ICU API allows the value of the exponent string to be the empty string. However, it turns out that this does \*not\* turn off exponent processing in ICU, it just sets the expected exponent character to be empty string! Which means a value such as

'1234+02' gets interpreted as '1234E+02' using ICU's lax parsing rules, instead of giving an error.

#### **7. DFDL support for transaction log data emitted by IBM 4690 point of sale terminals**

See email. These use a particular style of BCD, and allow variable length hexBinary strings. DFDL does not support this today.

#### **8. To help ensure compliance , provide error codes for all possible error scenarios**

This kind of thing is done by some of the W3C specs. If such an error code exists, and is output by DFDL processor implementation error messages, it allows TDML files for failure tests to be exchanged without modification. Need to decide on granularity of error codes.

#### **9. 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)  
Mike Beckerle

##### **Apologies**

Tim Kimber (IBM)  
Suman Kalia (IBM)

##### **Minutes**

The minutes cover the calls of 18th and 19th June. Minutes from 18th below.

#### **1. Daffodil Open Source Project**

To be discussed.

#### **2. Does DFDL need to support signed integer types with lengthUnits 'bits'?**

Spec only allows unsigned integers to have lengthUnits 'bits', but it is possible in C structures to declare bit fields for signed integers. They behave like two's complement. Agreed that DFDL should support this. Errata taken to permit lengthUnits 'bits' with signed integer types.

#### **3. Unsupported style of zoned decimal overpunching encountered**

On HP NonStop Tandem architecture, the negative sign is incorporated in the last byte of the number but using the highest bit (ie, 8) of the nibble, instead of the second highest bit (ie, 4). Agreed that DFDL should support this. Errata taken to add a new textZonedSignStyle enum 'asciiTandemModified'.

#### **4. Should parser check for parent 's terminating delimiters after parsing each child ?**

To be discussed.

#### **5. Agree on signature for new DFDL function to construct a DFDL string literal**

Agreed that making the DFDL function idempotent breaks a more common use case to the one that it fixes. The function will not be sensitive to DFDL string literals in the data . Steve will resend the mail with the function signature.

Should we also change DFDL function names to be hyphenated to match XPath's?

#### **6. ICU exponent symbol behaviour .**

To be discussed.

#### **7. DFDL support for transaction log data emitted by IBM 4690 point of sale terminals**

To be discussed.

#### **8. To help ensure compliance , provide error codes for all possible error scenarios**

To be discussed.

**Meeting closed**

**Next call**

#### **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: 176

#### **Actions raised at this meeting**

No	Action

#### **Current Actions :**

No	Action
123	<b>DFDL tutorial (Steve)</b> 13/10: Draft of first 3 chapters has been written and will be distributed to WG 10/11: Posted to grid forge here ( <a href="http://forge.gridforum.org/sf/go/doc16106?nav=1">http://forge.gridforum.org/sf/go/doc16106?nav=1</a> ), 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

01/12: Alan has started lesson 4 which covers fixed and variable fields and arrays.

08/12: Alan has almost completed lesson 4. Will send out for review.

15/12: First draft of lesson 4 is available for review. Alan to send to Bob and Joe.

22/12: Alan has distributed drafts for tutorials on Basic Structure and Optional/Repeating elements. Please review

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.

19/01: The tutorials for basic structure, optional/arrays and choices have been updated.

Please review. The tutorial for text elements should be available soon.

26/01: No comments received about 3 tutorials distributed last week. Alan is still working on Text representation.

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.

09/02: Steve had reviewed tutorials 3,4,5 and updated versions have been distributed. Joe reviewed lesson on text elements.

Main points. Using 'represented as text' is confusing. Examples are too cluttered. Suggest simple targeted examples but still build up to final complete schema

23/02: New versions distributed and Steve has commented.

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.

09/03: Alan distributed an update to the text tutorial. Please review.

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.

13/04: Steve is working on the text properties tutorial.

04/05: No progress

...

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.

28/06: On hold.

...

29/11: Tim offered to take a look at the next outstanding tutorials. Steve / Tim to discuss

6/12: No progress

...

10/01: No progress, offer from Mike to help. First step is to make any corrections due to errata.

17/01: No progress

24/01: No update

31/01: Daffodil project team will be working their way through the existing tutorials and reviewing

14/02: Daffodil team to start reviewing tutorials hopefully this Friday.

21/02: Moved to this coming Friday

28/02: No update

13/03: No progress

21/03: No progress from Daffodil team. IBMers are starting to use the tutorial and will feedback any comments.

28/03: No change

05/04: Steve will send Alan's two draft lessons on binary & text data to Mike to complete.

17/04: No progress

8/5: No update

	<p>...</p> <p><b>19/6: No progress</b></p>
140	<p><b>Spec issue: Parsing: 'missing' v 'empty', role of initiators , default values (All)</b></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> <p>1/11: Extra calls still ongoing</p> <p>8/11: Extra calls still going, when action 140 document complete will send to Steph for review</p> <p>15/11: Mike is verifying the action 140 conclusions by writing an algorithm in SCALA which can be ultimately be used in Daffodil</p> <p>22/11: Call to be held next week</p> <p>29/11: Next call 30/11</p> <p>6/12: Next call 7/12</p> <p>13/12: Had call earlier today, making good progress. Next call first week of Jan.</p> <p>10/01: Next call to be scheduled for Wed or Fri</p> <p>17/01: No call last week, call tomorrow</p> <p>24/01: Two calls held, next call Wed, looking at separator suppression</p> <p>31/01: Separator suppression discussions ongoing, proposal to rename policy and enums. Call later this week.</p> <p>14/02: Now looking at separator suppression and unparsing. Next call on Friday</p> <p>21/02: Separator suppression on parsing/unparsing matrix agreed.</p> <p>28/02: Two more calls this week</p> <p>13/03: Call tomorrow</p> <p>21/03: Two more calls held</p> <p>28/03: Hopefully all issues now addressed. Steve to start folding in action 140 document comments into the body of the document.</p> <p>05/04: Still with Steve. Noted that action 140 will not be in the next spec rev, likely the one after.</p> <p>17/04: No further progress</p> <p>8/5: Steve has started rewriting the action 140 document</p> <p>23/5: Steve continuing the work on the action 140 document. Empty/missing/defaults and Arrays have been rewritten. Separators not started. Please review.</p> <p>12/6: Steve will resend the latest action 140 document for review. Note use of 'missing representation' to describe zero length input data with same semantic as missing altogether.</p> <p><b>19/6: Latest action 140 resent. Separate call on Thursday this week.</b></p>
146	<p><b>Model NRL's bit-oriented format where complex repeating element carries a 'last' indicator (Steve, Mike)</b></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>

A more usable solution to the 'last indicator' problem is desirable though. Noted that this has a 'repeat..until' semantic.

16/08: Adam and Ryan to evaluate Steve/Tim proposal as it is potentially fragile relying as it does on backtracking.

23/08: Adam and 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.

4/10: Mike will contact NRL to see how things are progressing.

18/10: No update - Steve & Mike to chase

1/11: No update

8/11: Update on NRL via Mike - staffing the team to carry forward the DFDL work

15/11: No update

22/11: No update

29/11: NSA still expressing an interest in Daffodil

6/12: No update

13/12: NSA held a meeting with Steve & Mike to discuss their contributing to Daffodil through 2012

10/01: Will be added as an action item for the Daffodil Open Source Project

...

17/04: On hold

8/5: Discussed with NSA at CTDF follow-up meeting. This is the VMF format. They will actively look at this action.

23/5: Jeffrey Jacobs from NRL assigned to this action.

12/6: Expected to be looked at soon

**19/6: No update from Jeffrey**

161	<p><b>Fold errata into DFDL spec (Mike/Steve)</b></p> <p>17/01: Start with Steve updating spec to include his marked up changes</p> <p>...</p> <p>14/02: Steve has published errata document v8 which includes all resolved errata up until today, but has not yet started to fold into spec.</p> <p>21/02: No progress but Steve will start this week</p> <p>28/02: Steve has started on this. Editorial corrections made. Going through spec and flagging (using Word comments) where errata affect the spec (up to 2.16 so far). If the errata is simple then he is fixing the spec at the same time. The outcome of this first pass will be a spec in which a reader can see which sections are affected by errata. Then the second pass is to fold in the rest of the errata.</p> <p>13/03: Steve has completed first pass and sent spec to Mike. Mike has made his editorial changes and is applying errata. Steve to check formatting not scuppered by Mike's editor.</p> <p>21/03: Converted spec to docx format as this works better with Mike's editor, though Mike will switch to using Word for consistency. Mike folding in errata this week.</p> <p>Errata doc is due an update but that is on hold until Mike has finished, as each errata will be tagged as folded in or not.</p> <p>28/03: Steve has received updated spec from Mike. A few errata posed issues, Steve to resolve and respond to Mike. Mike to update spec and send back to Steve to review.</p> <p>05/04: Steve is reviewing Mike's updated spec and adding comments for issues. Should be complete next week. Need to decide on numbering of spec rev, Steve to seek guidance from OGF board. Still need to resolve errata 2.54 on strict v lax parsing.</p> <p>17/04: Steve has reviewed Mike's updated spec. Based on existing schemes for spec and errata, and taking into account OGF rules, the naming convention will be:</p> <ul style="list-style-type: none"> <li>- "DFDL spec 1.0.3 - GFD-P-R.174" (document gwdrp-dfdl-v1.0.3.doc)</li> <li>- "Errata for DFDL spec 1.0.3" (document ogf-dfdl-v1.0-Errata-xxx.doc)</li> <li>- "DFDL spec 1.0.4 - Draft (document gwdrp-dfdl-v1.0.4-Draft-xxx.doc)</li> </ul> <p>Note that xxx is the version number on GridForge.</p> <p>8/5: Steve has received updated spec from Mike and needs to review it. Steve has created draft errata v009.</p> <p>23/5: Steve to review Mike's spec. Steve to circulate error v009. Ideally errata 009 and action 140 should be included in next published spec, Mike to decide whether an intermediate spec at errata v008 level or errata v009 level is needed for Daffodil.</p>
-----	--

	12/6: Steve seeing formatting loss in Word 2003. Steve will install Word 2010 and see if that is the problem. Steve will resend errata v009 for review. Next step is to get action 140 incorporated into errata v010. <b>19/6: Steve has applied for license at IBM, but will also check from home PC.</b>
170	<b>Provide words for errata 2.54 on ICU lax/strict parsing of text number patterns (Richard - IBM)</b> 17/04: Need to define the basic (lax) parsing behaviour then define strict behaviour. 8/5: No progress. 23/5: Richard has started to create a set of tests to explore the behaviour when parsing 12/6: Continuing <b>19/6: Continuing</b>
171	<b>Decide on precise applicability of new encodingErrorPolicy property (All)</b> 23/5: To what stages of the parsing and unparsing processes does it apply, and to what types. 12/6: Tricky one to solve and could possibly defer to DFDL 2.0? Meanwhile see what MRM does. <b>19/6: Ask Tim to see what MRM does</b>
172	<b>Clarify how a DFDL string literal is matched against the data stream (Tim)</b> 23/5: Non-trivial algorithm, worth stating it in the spec. 12/6: No progress <b>19/6: No progress</b>
173	<b>Review Mike's DFDL string literal usage spreadsheet (All)</b> 23/5: May be found here: <a href="https://docs.google.com/spreadsheet/ccc?key=0AkXeiz4S7y0pdHpOdkJaWEtrS0kzWThEel9KamhGdkE#gid=0">https://docs.google.com/spreadsheet/ccc?key=0AkXeiz4S7y0pdHpOdkJaWEtrS0kzWThEel9KamhGdkE#gid=0</a> 12/6: No progress, review for next call please <b>19/6: Reviewed by Steve and spreadsheet updated during call. Key items from the review:</b> a) When nilKind is 'literalCharacter' or 'literalValue' and rep is binary, should we allow normal characters and DFDL character entities as well as raw bytes and %ES. Check with Tim if this causes problems for an implementation. This affects errata 20 which introduced the restriction based on result of call on 27th September 2011, although prior to that the restriction was even tighter for 'literalValue'. b) When property takes a logical value and is of type xs:string, is just allowing XML entities flexible enough? Two properties affected are nilValue and occursStopValue. Thought process had been that for nilValue just use 'literalValue' instead, and for occursStopValue use a terminator on a parent array element.

### Closed actions

No	Action
175	<b>Decide the fate of the overly strict parent /child alignment rule (All)</b> 12/6: Either drop altogether or limit to first child element., <b>19/6: Closed. Restriction will be dropped entirely.</b>

### Deferred actions

No	Action
129	<b>Press release to publicise DFDL (Steve)</b> 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	<b>Transformation of DFDL properties to a canonical form (Joe)</b>

	<p>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 gridforge site.</p> <p>15/12: Alan tested against test dfdl schema which worked correctly (after fixing some errors in the schema)</p> <p>22/12: no update</p> <p>12/01: Joe has some defects to fix before making available on gridforge.</p> <p>19/01: There is a difficult problem to solve before Joe make the style sheet public</p> <p>26/01: Working on problems</p> <p>02/02: no progress</p> <p>09/02: As it wasn't a simple as expected this will be treated as a low priority action</p> <p>23/02: Low priority</p> <p>09/03: Low priority</p> <p>30/03: Deferring for now</p>
066	<p><b>Investigate format for defining test cases (All)</b></p> <p>25/11: IBM to see if it is possible to publish its test case format.</p> <p>04/12: no update</p> <p>...</p> <p>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'</p> <p>24/03: No progress</p> <p>03/03: Discussions have been taking place on the subset of tests that will be provided.</p> <p>10/03: work is progressing</p> <p>17/03: work is progressing</p> <p>31/03: work is progressing</p> <p>14/04: And XML test case format has been defined and is being tested.</p> <p>21/04: Schema for TDML defined. Need to define how this and the test cases will be made public</p> <p>05/05: Work still progressing</p> <p>12/05: Work still progressing</p> <p>02/06: Work still progressing on technical and legal considerations</p> <p>...</p> <p>25/08: Will chase to allow Daffodil access to test cases. The WG should define how implementation confirm that they 'conform to DFDL v1'</p> <p>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.</p> <p>08/09: IBM still progressing</p> <p>15/09: IBM still progressing, expect tests to be available within a few weeks</p> <p>22/09: IBM still progressing, expect tests to be available within a few weeks</p> <p>29/09: Test cases are being prepared.</p> <p>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.</p> <p>13/10: Still progressing</p> <p>10/11: Legal issues cleared, IBM in process of collecting 100 example test cases, ideally ones that fit the 'extended conformance' of NCSA Daffodil</p> <p>17/11: Work is progressing on verifying the test cases. It should be possible to distribute to the WG in 2 weeks.</p> <p>24/11: About half the test cases have been completed and are being reviewed internally.</p> <p>01/12: Test cases should be available shortly</p> <p>08/12: The test cases are in internal IBM review. Probably need a bit of reorganising before publication</p> <p>Stephanie gave a brief overview of the format of the test cases.</p> <p>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.</p> <p>22/12: Test cases were sent to Joe for initial testing which found some problems in the Daffodil parser</p> <p>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.</p> <p>19/01: There is currently no resource available in IBM to make more tests available. IBM to</p>

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
<b>043</b>	Track errata list for 1.0 of the spec. <a href="http://forge.gridforum.org/sf/go/doc16280?nav=1">http://forge.gridforum.org/sf/go/doc16280?nav=1</a>	Steve	N/A	Draft 0008 on grid forge.
<b>044</b>	Incorporate errata list into DFDL spec.	Steve/Mike	N/A	Draft merged document.