

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	17-Jul-12 (Tues)
Meeting Time	15:00 - 16:00

Created by Steve Hanson on 09-Mar-11

Last Modified by Steve Hanson on 17-Jul-12

OGF DFDL Working Group Call , 17 July 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. Nested discriminators

Clarify how a discriminator relates to a point of uncertainty .

3. Clarify behaviour for numbers when excess precision supplied in infoset and rounding switched off

Text: Spec says that when rounding is not enabled (either via property or pattern) then no rounding takes place.

Should say that any number out of range causes a processing error (this is ICU's behaviour).

Binary: Spec does not say what happens. Should be a processing error.

4. 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)

Tim Kimber (IBM)

Suman Kalia (IBM)

Apologies

Mike Beckerle (IBM)

Minutes

1. Daffodil Open Source Project

[No update.](#)

2. Nested discriminators

Feeling is that a discriminator is not tied physically to its nearest point of uncertainty , but when evaluated the result applies to the nearest active point of uncertainty in scope. Needs some more discussion so new **action 180** raised.

3. Clarify behaviour for numbers when excess precision supplied in infoset and rounding switched off

Agreed that when rounding is not enabled (either via property or pattern) and excess precision is supplied in the infoset then it is processing error. Applies to text numbers (where rounding is an option) and binary numbers (where rounding is not currently an option). Matches ICU behaviour. **Errata taken**.

Meeting closed

15:40 UK

Next call

TBD as Steve is away next Tuesday.

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: 181

Actions raised at this meeting

No	Action
180	Clarify the point of uncertainty to which a discriminator applies 17/7: Nearest physically or nearest active at runtime?

Current Actions :

No	Action
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>...</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>29/11: Tim offered to take a look at the next outstanding tutorials. Steve / Tim to discuss</p> <p>6/12: No progress</p> <p>...</p> <p>10/01: No progress, offer from Mike to help. First step is to make any corrections due to errata.</p> <p>17/01: No progress</p> <p>24/01: No update</p> <p>31/01: Daffodil project team will be working their way through the existing tutorials and reviewing</p>

	<p>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 ... 17/7: No progress</p>
140	<p>Spec issue: Parsing: 'missing' v 'empty', role of initiators , default values (All) 01/06: See minutes. 08/06: Still under discussion. Tim has sent Mike a selection of data formats to guide the discussion. 15/06: Not discussed - an extra call has been scheduled to go through this. 28/06: A series of extra calls are being held between Mike, Steve, Tim and Steph. 05/07: Next extra call is Wed 6th July - Steve to send invite 12/07: Two more calls held. Next call is Wed 13th July. 19/07: More calls held, next call is Fri 22nd July. 26/07: More calls held, good progress 16/08: Steve will set up next call for when Tim has returned from holiday 23/08: Two more calls scheduled for this week, remaining issues: separator suppression, empty strings, sparse arrays (see action 136) 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. 20/09: Steve has summarised where we are with action 140, which Tim and Mike have reviewed. 2 hours call planned for Thursday. 27/09: Calls held, progressing the separator suppression behaviour 4/10: More calls held, progressing separator suppression, sparse arrays (see action 136) and emptyValueDelimiterPolicy behaviour 18/10: All issues now in a single document, call held earlier today. Next call Thursday. 1/11: Extra calls still ongoing 8/11: Extra calls still going, when action 140 document complete will send to Steph for review 15/11: Mike is verifying the action 140 conclusions by writing an algorithm in SCALA which can be ultimately be used in Daffodil 22/11: Call to be held next week 29/11: Next call 30/11 6/12: Next call 7/12 13/12: Had call earlier today, making good progress. Next call first week of Jan. 10/01: Next call to be scheduled for Wed or Fri 17/01: No call last week, call tomorrow 24/01: Two calls held, next call Wed, looking at separator suppression 31/01: Separator suppression discussions ongoing, proposal to rename policy and enums. Call later this week. 14/02: Now looking at separator suppression and unparsing. Next call on Friday 21/02: Separator suppression on parsing/unparsing matrix agreed. 28/02: Two more calls this week 13/03: Call tomorrow 21/03: Two more calls held 28/03: Hopefully all issues now addressed. Steve to start folding in action 140 document comments into the body of the document. 05/04: Still with Steve. Noted that action 140 will not be in the next spec rev, likely the one after. 17/04: No further progress 8/5: Steve has started rewriting the action 140 document 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>19/6: Latest action 140 resent. Separate call on Thursday this week.</p> <p>26/6: Call held, revisions need to be made before action 140 can be circulated more widely. Next call will be Tues 3rd July.</p> <p>3/7: Steve not found time to update action 140 doc, call postponed to 10th July</p> <p>10/7: Call held, document still being refined. Spin off action 179 to sort out use of term 'representation' in spec and grammar.</p> <p>17/7: No progress on core action 140.</p>
<p>146</p>	<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 and Ryan to evaluate Steve/Tim proposal as it is potentially fragile relying as it does on backtracking.</p> <p>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.</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> <p>1/11: No update</p> <p>8/11: Update on NRL via Mike - staffing the team to carry forward the DFDL work</p> <p>15/11: No update</p> <p>22/11: No update</p> <p>29/11: NSA still expressing an interest in Daffodil</p> <p>6/12: No update</p> <p>13/12: NSA held a meeting with Steve & Mike to discuss their contributing to Daffodil through 2012</p> <p>10/01: Will be added as an action item for the Daffodil Open Source Project</p> <p>...</p> <p>17/04: On hold</p> <p>8/5: Discussed with NSA at CTFD follow-up meeting. This is the VMF format. They will actively look at this action.</p> <p>23/5: Jeffrey Jacobs from NRL assigned to this action.</p> <p>12/6: Expected to be looked at soon</p> <p>19/6: No update from Jeffrey</p> <p>26/6: No update from Jeffrey. Also looks like ISO 8583:2003 has the same kind of continuation indicator.</p> <p>3/7: No update</p> <p>10/7: Jeffrey has not looked at the format in detail but thinks that support for continuation indicators is a good thing to have anyway.</p> <p>Want to close off this action so we need to decide whether DFDL 1.0 needs to support this in a better way than already demonstrated.</p> <p>Mike to check with Jeffrey on Daffodil status call today.</p> <p>17/7: Received email from Jeffrey, He has looked at the current proposal. Hoping that occursCount expression can handle the repeat indicator.</p>
<p>161</p>	<p>Fold errata into DFDL spec (Mike/Steve)</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</p>

	<p>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. 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. 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. 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. 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. 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. 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: - "DFDL spec 1.0.3 - GFD-P-R.174" (document gwdrp-dfdl-v1.0.3.doc) - "Errata for DFDL spec 1.0.3" (document ogf-dfdl-v1.0-Errata-xxx.doc) - "DFDL spec 1.0.4 - Draft (document gwdrp-dfdl-v1.0.4-Draft-xxx.doc) Note that xxx is the version number on GridForge. 8/5: Steve has received updated spec from Mike and needs to review it. Steve has created draft errata v009. 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. 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. 19/6: Steve has applied for license at IBM, but will also check from home PC. 26/6: Steve now has Word 2010 installed. Steve has published errata v009 and v010 on GridForge. Latest internal spec draft is at errata v008. Mike to consider updating the spec to include errata v009 and v010 before action 140 incorporated. Naming convention for spec drafts is not as stated above, but is now gwdrp-dfdl-v1.0.3-xxx.yy where xxx is Errata document version and yy is local draft. 3/7: Mike folding errata 10 document into spec, close to completion 10/7: Mike has sent updated spec for review. 17/7: Steve has not yet reviewed the updated spec</p>
170	<p>Provide words for errata 2.54 on ICU lax/strict parsing of text number patterns (Richard - IBM) 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 .. 17/7: Continuing</p>
172	<p>Clarify how a DFDL string literal is matched against the data stream (Tim) 23/5: Non-trivial algorithm, worth stating it in the spec. .. 17/7: No progress</p>
178	<p>OGF migration from SourceForge to Redmine (Steve) 10/7: Opened to track any work arising 17/7: No action so far.</p>
179	<p>Sort out use of 'representation' in spec and grammar 10/7: Spun off from action 140 17/7: Mike updated the grammar, Steve reviewed and revised. Team to review for next call.</p>

Closed actions

No	Action
176	Word 2010 export CSV format (Suman)

26/6: Ensure that this can be modelled using DFDL.
 3/7: No update.
 10/7: Suman will try and look at this this week.
 17/7: **Closed**. Tables exported from Word direct to CSV are generated with quotes around strings, but do not double up any contained quotes. This is not accordance with CSV rules, and from an escaping perspective it is broken. The recommendation seems to be to export the table to Excel first, then export as CSV, and that will apply correct escaping rules. Not worth adding anything specific to DFDL's delimited scanning to handle this broken format, could probably parse it ok using lengthKind 'pattern'.

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 priority 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</p>

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 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.

...

10/07/2012: Discussed schemes to create interchangeable tests. Ideally need a DFDL defined error code per failure, in conjunction with specific inserts.

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 010 on grid forge.
044	Incorporate errata list into DFDL spec.	Steve/Mike	N/A	Draft merged document.

