

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	21-Feb-12 (Tues)
Meeting Time	15:00 - 16:00

Created by Steve Hanson on 09-Mar-11

Last Modified by Steve Hanson on 21-Feb-12

OGF DFDL Working Group Call , 21 February 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 update.

2. Escape schemes and nil /empty processing

Ensure the spec behaves correctly when block style escaping is used with empty strings and with nil values .

Last week we said:

a) *empty*

Unparsing: Empty string will only be escaped if generateEscapeBlock is 'always'.

Parsing: If escapeBlockStart/End characters are present then the content region is not empty.

This will not cause a default value to be applied, the value is taken as empty string.

b) *nil value*

Parsing: Escape scheme applied before nil processing, so can't use escaping to switch off nil.

Unparsing: Escape scheme applied after nil processing.

Steve has looked at the property precedence section to see if it needs changing.

3. Clarify behaviour of textNumberPattern when using V /P and # symbols

IBM implementation team are questioning whether patterns like #V#, PP### and ###PP make sense.

V and P are DFDL extensions to the ICU pattern language to mimic COBOL and other formats that can not express a physical decimal point.

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)

Mike Beckerle

Alonza Mumford (NSA)

Tim Kimber (IBM)
Richard Schofield (IBM)

Apologies

Minutes

1. Daffodil Open Source Project

Many of the team received an education session on IBM DFDL from IBM.
Aiming to have something concrete for the NSA conference at end April.

2. Escape schemes and nil /empty processing

No progress on checking the property precedence, leave for next week

3. Clarify behaviour of textNumberPattern when using V /P and # symbols

Discussed a couple of alternatives:

a) Disallow use of # with V and P.

Conservative option. Assumes V and P used primarily with fixed length data.

Makes that you have to handle variable length numbers with VDP by using padding.

b) Allow unambiguous combinations of # and V and P only.

V: Must have only 0 symbols to the right of V

P (left): Must have as many 0 symbols adjacent to P as there are Ps.

P (right): No limitations

Decision was leaning towards a) but Steve wanted to check with Steph whether this would impact industry formats like SWIFT, **Action 167** raised.

4. Establish large file requirement from NSA

Multi GB file that contains several large records which each comprise fields and BLOB.

Alonza will talk to Boyd, ideally will supply a DFDL schema illustrating the model.

Meeting closed

Next call

Tues 28th Feb 2012 @ 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: 167

Actions raised at this meeting

No	Action
167	Establish SWIFT etc use of numbers with virtual decimal points (Steve) 21/2: Then we can decide what to do with patterns that contain #, V, P

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 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 be updated. Please review. The tutorial for text elements should be available soon. 26/01: No comments received about 3 tuorials 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
18/05: No progress
01/06: No progress
08/06: 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](#)

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
 18/10: Steve has met with legal, can publish the xsds after tidy up. Will create OGF info document.
 1/11: No further progress.
 ...
 29/11: Steve has a version of the default format available, waiting for one last IBM legal procedure.
 6/12: Waiting for one last IBM legal procedure.
 13/12: Still waiting for one last IBM legal procedure. OK to review iwhat we have so far though, so Steve sent out an xsd containing a single defineFormat and a single defineEscapeScheme.
 The spec in section 19 makes it clear that any such xsds are essentially examples, so this xsd will be treated as such. It will be in a 'DFDL example' namespace and we will publish as OGF information document on GridForge.
 Reviewed the properties and changed occursCountKind to 'parsed'. (Noted that 'implicit' is better but that is subject of an errata not yet folded back into the spec).
 Stayed with lengthKind 'delimited' with a comma separator so that the xsd would actually model a real single record CSV stream.
 Steve to improve the example usage section in the prolog to include namespaces and example data, and run some real CSV tests.
 10/01: IBM legal process complete. Steve has updated the xsd as above and used it to model CSV data. Steve to formalise as an OGF document.
 17/01: Steve has created an OGF document and mailed it to the WG, but noted that it has both IBM and OGF copyright. Steve to get advice on how to proceed from OGF editors.
 24/01: Mail sent to OGF, no reply yet
 31/01: Reply received from OGF, problem being discussed
 14/02: Update received from OGF, Steve to digest and work out how to publish.
 21/02: [Steve still in dialog with OGF, document sent to OGF board to approval](#)

140

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

	<p>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.</p>
145	<p>Provide a 'dispatch' way of discriminating a choice for better performance of the envelope/payload use case (Steve, Mike, Suman) 12/7: See minutes. Need to choose a proposal and flesh out. 19/07: Waiting for proposals 26/07: Waiting for proposals 16/08: Waiting for proposals. Suman added to action. ... 1/11: Steve to send a proposal ... 21/02: Waiting for Steve to find time to do this</p>
146	<p>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 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 ... 21/02: On hold</p>
159	<p>Property documentFinalTerminatorCanBeMissing not flexible enough ? (Steve) 10/01: Do we need to upgrade this property to handle formats where the terminator is not allowed on output but tolerated on input? ... 21/02: With Steve to investigate further</p>
161	<p>Fold errata into DFDL spec (Mike/Steve) 17/01: Start with Steve updating spec to include his marked up changes ... 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. 21/02: No progress but Steve will start this week</p>

Closed actions

No	Action
165	<p>Issues with using lengthKind endOfParent (Team)</p> <p>31/01: Agree on list of restrictions and best way to model a single record file with a BLOB on the end.</p> <p>14/02: Team to think on Mike's proposal which allows a global element to carry lengthKind 'endOfParent' even when the global element is used as the root, and for this to mean that the element ends when the data runs out.</p> <p>21/02: Closed: Steve resent the email that lists the restrictions. Agreed on list. Agreed that endOfParent allowed on complex root elements. Existing restrictions in spec will be augmented with new ones. Errata taken.</p>

Deferred actions

No	Action
129	<p>Press release to publicise DFDL (Steve)</p> <p>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.</p> <p>17/11: no progress</p> <p>...</p> <p>08/12: Still no response from IBM press office</p> <p>15/12: no progress</p> <p>....</p> <p>09/03: No progress</p> <p>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)</p> <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 gidforge 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 exoected this will be treated as a low priority action</p> <p>23/02: Low prioity</p> <p>09/03: Low priority</p> <p>30/03: Deferring for now</p>
066	<p>Investigate format for defining test cases (All)</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>

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 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 0007 on grid forge.
044	Incorporate errata list into DFDL spec.	Steve/Mike	TBD	