TeamRoom Plus

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

Meeting about Meetings\OGF

Project	DFDL 1.0
Meeting Date	18-Sep-12 (Tues)
Meeting Time	15:00 - 16:00

Created by Steve Hanson on 09-Mar-11
Last Modified by Steve Hanson on 18-Sep-12

OGF DFDL Working Group Call , 18 September 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. Clarification on aspects of variables behaviour   (more)
Section 7.7 says "*A variable can have any simple type*". Clarify whether this includes:
- Simple types that are not in the DFDL schema subset (suggest no)
- User-defined simple type restrictions (suggest yes)

Section 7.7 says "*The scope of a variable name covers the entire schema in which it is
defined and any imported or included schema* " as per errata 2.12. Are these words clear enough?

3. Should DFDL have built in support for IBM   4690 TLOG data ?
See email from Steve, which explains the IBM TLOG packed field format and proposes how it could
be supported by DFDL spec.

4. Use of the terms 'optional' and 'required'.
Ensure that the use of these terms in the spec is unambiguous or qualified.   See e-mail from Mike.

5. Spec description of text number rounding is not how ICU behaves   .
It looks like there is a bug in the ICU4J documentation. It implies that no rounding takes place if
the rounding increment is 0.0 (or equivalently only '0' digits appear in the pattern). The spec behaviour
was based on this wording. Tests show that ICU4C and ICU4J both treat 0.0 as 'round to the
width of the pattern'. The way to turn off rounding altogether is to be to use rounding mode 'roundUnnecessary'
which is currently an enum not exposed by DFDL. Interestingly, the IBM DFDL model has added enum 'none'
into its definition of dfdl:roundingMode. IBM is clarifying the correct behaviour with ICU, and will propose
a spec change depending on the outcome.

6. Timing of asserts and discriminator evaluation
Spec currently says "*Any element referred to by the expression must have already been processed or is a
descendent of this element. The expression must have been evaluated by the time this element and it
descendents have been processed.* " plus additionally for discriminators *"or when a processing error occurs
when processing this element or its descendents* ." Descendent references are useful as they permit a
discriminator to be lexically associated close to the point of uncertainty . However this poses an issue about
when to evaluate the expression. The parser needs to evaluate the expression without being fooled about

whether something temporarily doesn't exist or not. Discuss.

7. AOB

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

Apologies

Minutes

1. Daffodil Open Source Project

2. Clarification on aspects of variables behaviour   (more)

3. Should DFDL have built in support for IBM   4690 TLOG data ?

4. Use of the terms 'optional' and 'required'.

5. Spec description of text number rounding is not how ICU behaves   .

6. Timing of asserts and discriminator evaluation

7. AOB

Meeting closed

Next call
Tues 25th Sept 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: **185**

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 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 respresentations. 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 ... 4/9: No progress: 11/9: IBM DFDL infocenter will start to reference these directly before the end of the year, so they need updating soon. 18/9:
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

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.

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.

19/6: Latest action 140 resent. Separate call on Thursday this week.

26/6: Call held, revisions need to be made before action 140 can be circulated more widely. Next call will be Tues 3rd July.

3/7: Steve not found time to update action 140 doc, call postponed to 10th July

10/7: Call held, document still being refined. Spin off **action 179** to sort out use of term 'representation' in spec and grammar.

17/7: No progress on core action 140.

25/7: No progress - Steve will aim to finish refinement before Aug 16

31/7: No progress

7/8: No progress

14/8: Steve has started to create v16 of the action 140 document and will mail it to Mike & Tim this week

4/9: v16 mailed to Mike & Tim - some issues noted in the email - Steve will set up a separate call

11/9: Separate call held. Re-examining the role of minOccurs for occursCountKind 'parsed', 'expression', 'stopValue'. For these occursCountKinds, where the occurrences are extracted without reference to minOccurs, it seems more natural that a minOccurs violation is not a processing error but instead it is just a validation error. Need to evaluate the knock-on effect of this proposal - it potentially affects points of uncertainty, default values, use of terms required & optional, and separator suppression.

18/9: Tim and Steve have worked through the proposal, it looks good in principle. Steve to update action 140 document and see if anything problematic surfaces.

| 161 | 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
	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.
	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-rxxx.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
	25/7: Steve has not yet reviewed the updated spec - aim to do so before Aug 16
	31/7: No progress
	7/8: Steve has reviewed most of the spec, and will complete by end of the week. Spec updated.
	14/8: Meeting held to review remaining open comments in latest spec, Mike will update and mail out
	4/9: Mike updated the spec and mailed it to Steve. Steve to review.
	11/9: Still with Steve to review.
	18/9:
170	**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
	..
	25/7: Continuing.
	31/7: Continuing, Richard expects to have something next week. Steve has asked ICU for

similar statement of behaviour for lax/strict calendar parsing.

...
14/8: On hold while Richard on holiday.
4/9: Steve to talk to Richard .
11/9: See email from Richard. In summary:
Strict mode (ie, base behaviour):
- Digits 0-9, decimal separator will be parsed successfully regardless of the pattern or strict/lenient parsing mode
- Grouping separators are only permitted in the data, if specified in the pattern
Additionally in lenient mode, the following do not cause a failure:
- Leading or doubled grouping separators
- Groups of incorrect length when grouping is used
- Grouping separators used in numbers followed by exponents
- Quoted characters in the pattern must be present in the data.
While the base behaviour is not ideal for DFDL, agreed that for now DFDL will adopt the ICU rules on the grounds of ease of implementation. **Errata taken** .
As a follow-on , ICU are talking about introducing an enum to control strict/lenient modes. IBM team will discuss with ICU the provision of a mode that corresponds to ideal DFDL behaviour, which is to honour the pattern as close as possible .
DFDL textNumberCheckPolicy can then be extended with a new enum to match.
18/9: Awaiting ICU outcome

172	**Clarify how a DFDL string literal is matched against the data stream (Tim)**
	23/5: Non-trivial algorithm, worth stating it in the spec.
	..
	25/7: No progress.
	31/7: Tim has been making notes but nothing written up formally. Will include treatment of %WSP*;
	...
	18/9: No further progress

178	**OGF migration from SourceForge to Redmine (Steve)**
	10/7: Opened to track any work arising
	17/7: No action so far.
	25/7: No action so far.
	31/7: Steve has registered at the Redmine site. There is a skeletal section for DFDL WG. Admin for mailing list has been automatically migrated . Nothing for documents, trackers, tasks, source code.Steve has emailed the site admin for migration instructions . Need to decide what artifacts we want to migrate - certainly documents and source code.
	7/8: No reply received, Steve will chase.
	14/8: Still no reply received - probably on holiday
	4/9: Reply received. GridForge to be read-only after OGF 36. OGF have offered to help migrate DFDL-WG content to RedMine. Steve has sent list of content.
	...
	18/9: No update from OGF yet

183	**Clarify unparsing behaviour when lengthKind is 'explicit ' and length is an expression (Steve)**
	4/9: We know what the DFDL schema looks like for this, need to state implementation behaviour for various sub-scenarios
	11/9: See email from Steve. It contains 4 scenarios to illustrate the issues around unparsing 'explicit' when length is an expression.
	Mike proposed that when there is an expression, the data is variable length and therefore no padding should take place at all. This removes the chicken-and-egg issue.
	Tim suggested using a new 'expression' enum but from where we are today, it is less disruptive to just look for expression syntax {}.
	Steve noted that we would still have to still pad to minimum lengths for text, or implicit lengths for binary, so same rules as for 'prefixed'.
	Mike had a scenario where the length to pad to was dependent on the length of the data (eg, 1,2,3 => pad to 4, 5,6,7 => pad to 8).
	Need to make sure that the proposal does not scupper this scenario, Steve thought that it might do.
	18/9:

No	
184	**Can variables be simplified in any way ? (Mike)** 11/9: Are there any usability improvements that can be made to the current set of annotations? It looks like the DFDL expression production rules allow a variable as a path segment, is that intentional? 18/9: We need the three annotations. Suggestion made to add a 'variableKind' attribute to limit how a variable is used (eg, only via newVariableInstance). Although DFDL expression production rules appears to allow a variable as a path segment, that is not supported by DFDL, which can only return simple values. XPath does not allow variables to return simple values in paths, only nodes.

Closed actions

No	Action

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 abot 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.