TeamRoom Plus

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

Meeting about Meetings\OGF

Project	DFDL 1.0
Meeting Date	17-Apr-12 (Tues)
Meeting Time	15:00 - 16:00

Created by Steve Hanson on 09-Mar-11
Last Modified by Steve Hanson on 20-Apr-12

OGF DFDL Working Group Calls , 17 & 20 April 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. Strict v lax parsing
Errata 2.54 has revealed issues around the specification of the ICU lax and strict parsing behaviour
for text numbers. Need to define the basic (lax) parsing behaviour then define strict behaviour.

3. DFDL and recursive models
The spec currently disallows recursion primarily on simplicity grounds . Is this still ok for DFDL 1.0 ?

4. ESA DFDL-based application
The European Space Agency has released an application for visualizing binary satellite data
(compliant to CCSDS standard) which based on DFDL. It is called S2G Data Viewer and can
be freely downloaded from http://eop-cfi.esa.int/S2G/s2g.html. It uses a small subset of
DFDL, but it is not compliant with the DFDL spec in several respects .
Steve has been in dialog with the ESA and they will attend the DFDL WG call on May 8th to discuss.

5. Choosing the correct branch when serializing .
See email. Because of DFDL allowing infoset elements to be missing when parsing, DFDL needs to be
clear on the rules used to match a sparse infoset to a model. Specifically a problem when the children of
the choice are model groups.

6. Errata 2.22 revisited
Errata 2.22 says that textStandardDecimalSeparator must be set if pattern contains a '.' but it must also
be set if the pattern contains 'E' or '@' as these on their own can cause a decimal separator to be output
(due to use of significant digits logic). Errata needs to be updated to reflect this.

7. Errata 2.70 revisited
See email. Clarify wording around delimiters and ordering longest first / greedy algorithm where variable
length DFDL entity classes are involved.

8. Errata 3.13 revisited
Clarify whether the new encodingErrorPolicy property applies when doing encoding conversions
as part of matching delimiters, or padding/trimming, or nil comparison (literal and logical) or

text boolean true/false rep.

9. Clarifications around expression language and DFDL string literals
Need to make several things clearer:
- Ensure that everywhere that accepts an expression states exactly what it is expected to return
- Use XPath constructors or the correct literal values to ensure the returned value is of the correct type
- What is returned follows XPath 2.0 rules - note that this is not the same as xs:default / xs:fixed
- Should we allow any auto-casting a) for simple type restrictions, b) for wider mis-matches
- If not expecting expression to return a DFDL string literal, the returned value is not treated as such
- If expecting expression to return a DFDL string literal, the returned value is treated as such
- Provide a DFDL string literal constructor function so that values from data stream are DFDL string literal compliant ?
- <u>Within</u> an expression, a string is never interpreted as a DFDL string literal

10. Clarify the description of dfdl :property and other DFDL functions
The spec is not clear as to whether dfdl:property returns the actual coded property value, the property value with expressions evaluated, or the property value with DFDL entities replaced, or a combination of the last two.
In fact the spec does not provide sufficient description for any of the DFDL functions.

12. DFDL string literal usage diagram
Review Mike's diagram

13. Clarify that DFDL regular expressions do not interpret DFDL entities

14. AOB
DFDL WG call Tues 24th April & Tues 1st May. Steve is away.

Minutes

<p style="text-align:center">**Meeting Minutes**</p>

Reflect on your meeting as you record all topics and issues discussed , and any tabled conversations. What went well, or what would you do differently next time? Document those so others can take advantage of your learning .

Attendees
Steve Hanson (IBM)
Suman Kalia (IBM)
Mike Beckerle
Tim Kimber (IBM)

Apologies

Minutes

1. Daffodil Open Source Project
No update.

2. Strict v lax parsing
New **action 170** to create words that be can be used for errata 2.54 and spec.
Need to define the basic (lax) parsing behaviour then define strict behaviour.

3. DFDL and recursive models
Agreed that DFDL 1.0 still disallows recursion.

4. ESA DFDL-based application
Will be discussed with ESA on 8th May call.

5. Choosing the correct branch when serializing .
This is only an issue when the children of a choice are sequences or choices. Decided that the first element in the infoset will be used to search the choice branches in order. If the element occurs in that branch then that will be chosen, even if there is a 'better' match in another branch. This keeps the rule

simple. No backtracking, so if a branch subsequently fails, it is a hard failure. Modeler can always wrap the sequence/choice in an element to resolve any problems with unwanted matches. Errata taken.

6. Errata 2.22 revisited
Errata taken to update 2.22. to say that textStandardDecimalSeparator must be set if pattern contains '.' or 'E' or '@' as all of these can cause a decimal separator to be output.

7. Errata 2.70 revisited
Errata taken. Mike will clarify wording to cover DFDL entity classes.

Meeting closed
This was a two part meeting held on 17th & 20th April.

Next call
24th April call cancelled as Mike and Steve are at US NSA CTDF conference.
1st May call moved to 2nd May.

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: **171**

Actions raised at this meeting

No	Action
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.

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 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
		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
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
146	Model NRL 's bit-oriented format where complex repeating element carries a 'last' indicator (Steve, Mike)	
		19/07: NRL to send xsd and example of format.

	26/07: Steve/Tim put forward a proposal that used a discriminator that looked at the previous item's repeatBit in the array and failed if it was 0. This looks like it will work. Mike to add the correct DFDL annotations to take into account both repeatBit and presentBit in a single discriminator, and auto-set them using outputValueCalc. A more usable solution to the 'last indicator' problem is desirable though. Noted that this has a 'repeat..until' semantic. 16/08: Adam 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
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.

Closed actions

No	Action
145	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 ... 21/03: Steve has sent a proposal. Mike has sent a counter proposal. Steve to respond. 28/03: Steve has sent a revised proposal. Review for discussion next week. Ensure proposal handles Mike's scenario where tag value to branch mapping is not 1-1. 05/04: Discussed Mike's review comments and Suman's concerns. Agreed that name should be elementID, should be a single DFDL String Literal value, and that matching of choiceBranchRef expression result should only be against elementID to avoid QName v String confusion. Steve to recirculate with a schema example. 17/04: **Closed**. Discussion on whether the choiceBranchRef expression should retiurn xs:string or something else. Agreed on xs:string. Discussed whether elemenID should be a pure xs:string or a DFDL String Literal. For consistency with other DFDL properties it should be a string literal, but raw byte entities and character classes should be disallowed to avoid complications. Discussed scope of uniqueness of elementIDs. Agreed that uniqueness is both local to a choice, and across all global elements in the same namespace (the latter is not strictly needed right now but accommodates any future addition of xs:any). Agreed that elementID should be on global element, local element, and element ref (in which case it overrides any elementId on the global element, which is ok as the property does not follow the usual scoping rules). Errata taken.
169	**Establish nature of US NSA conference DFDL tutorial (Alonza)** 28/03: Need to know whether this is a hands-on lab or a watch-me tutorial. Mike to email Alonza. 05/04: No facilities in venue for hands-on lab, so will be a tutorial. Next step is to get a couple of example formats from Alonza. 17/04: **Closed**. Looks like USMTF is a good basis for the tutorial. Steve has sent tutorial slides to Mike, and will send schemas.

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.

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 0008 on grid forge.
044	Incorporate errata list into DFDL spec.	Steve/Mike	N/A	Draft merged document.