DFDL WG Call Agenda

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

Meeting about Meetings\OGF

Project	**DFDL 1.0**
Meeting Date	**23-Sep-14 (Tues)**
Meeting Time	**14:00 - 15:00**

Created by Steve Hanson on 09-Mar-11
Last Modified by Steve Hanson on 22-Sep-14

OGF DFDL Working Group Call , 23 Sep 2014

Agenda

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

1. Daffodil Open Source Project
Status update.

2. Fixed length elements with initiators
Deferred from last call.
Steve has noticed a possible catch 22 with the setting of emptyValueDelimiterPolicy in this scenario . See email.

3. XPaths: OR expression evaluation
Should DFDL allow lazy evaluation (working left to right), or insist on full evaluation?
Daffodil currently implements lazy evaluation via its use of Saxon . See email.

4. XPaths: Namespaces of element names in the path
Clarify how path steps are to be interpreted taking into account elementFormDefault 'qualified'/'unqualified' and
target/notarget namespaces. See email.

5. 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

Apologies

IPR Statement
"I acknowledge that participation in this meeting is subject to the OGF Intellectual Property Policy ."

Minutes

Meeting closed

Next regular call
TBD

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

Subject	Document Type	Created	Modified

Next action: **272**

Actions raised at this meeting

No	Action

Current Actions :

No	Action
228	**Review set of tutorial lessons (All)** 17/9: Lesson 1 proposes a set of lessons, needs reviewing as over 2 years old. ... 22/10: No progress

31/10: Becoming a focus for Tresys. Steve to send his 'Modeling Data Formats using DFDL' powerpoint.
...
19/11: No further progress
26/11: Possibility of help from MITRE high-school student, and from Marisa at IBM.
...
11/3: No further progress
25/3: MITRE have produced a couple of new tutorials under the guidance of James Gariss. Jonathan to forward for review.
Mike observed that an html tutorial could be generated from a tdml file using XSLT.
11/4: Not discussed
15/4: Jonathan will send 4 new mini-tutorials. Need to figure out best way to incorporate into the tutorial structure.
29/4: Tutorials received. Mark has taken a quick read. Mark & Steve to review and report back.
6/5: Still with Mark and Steve
20/5: Mark has reviewed. Will ask IBM information development to recommend a way to portray the existing and new lessons, preferably web-based. Find somewhere to host them. OGF? GitHub? developerWorks? NCSA?
3/6: Steve has also reviewed.
...
17/6: No further progress on tutorials. Tim is looking into the creation of some DFDL how-to videos using the IBM Integration Studio.
...
23/9: No further progress

| 248 | **Discriminators and potential points of uncertainty   (Steve)**
28/1: Steve to write up a proposal to prevent a discriminator from behaving in a non-obvious manner when used with a potential point of uncertainty that turns out not to be an actual point of uncertainty.
5/2: Steve sent an email to check whether choice branches, unordered elements and floating elements should always be actual points of uncertainty, as there are times when there is no uncertainty, eg, last choice branch; all floating elements found. It was decided that they are always actual points of uncertainty. To do otherwise will complicate implementations and result in fragile schemas. Steve will proceed with the proposal on that basis.
...
25/3: No further progress
11/4: Proposal sent to mailing list by Steve. Concern that having a potential PoU that in practice can never be an actual PoU is counter intuitive and we are better off saying that for certain occursCountKinds there is no potential PoU. The behaviour is therefore the same as for scalar elements. Means that occursCountKind 'fixed' and occursCountKind 'implicit' with minOccurs=maxOccurs behave differently wrt to discriminators. Steve will reword the proposal accordingly.
...
29/4: No further progress
6/5: Steve came to reword the proposal to say that for certain occursCountKinds there is no potential PoU, but it raised an issue. Steve has resent the original proposal with responses to Tim's questions. It is clear that a discriminator inside an array can not leak outside the array because it is evaluated for each occurrence. But should that be expressed by saying that a) all arrays are potential PoUs and a discriminator can't leak outside a PoU, or b) only some arrays are potential PoUs and a discriminator can't leak outside a PoU or an array. Please can WG members review the email and have a position on the wording.
20/5: Tim has reviewed, back with Steve
...
3/6: No progress
10/6: Not complete. Decided that next published specification would not include this.
...
26/8: No progress |

	2/9: No concrete progress although Steve came across this exact scenario when modelling NACHA Addenda records, and used asserts instead of discriminators to side-step the issue. ... 23/9: No further progress
250	**Standardise on a single tdml format for DFDL tests (All)** 5/2: Steve has requested permission for IBM to view / use the Daffodil tdml files, as a precursor to trying to standardise on a common tdml format. Was formerly part of action 066. ... 18/2: No further progress 11/3: Mike and Steve discussing the best way to share and cooperate on tdml format. 25/3: Discussed the creation of an OGF document that will own and define a standardised tdml format. 11/4: Proposal is for the OGF document to define a tdml format without Tresys or IBM copyright statement. 15/4: Draft document on Redmine ... 6/5: No further progress 20/5: Mark has read through the document. Particularly concerned with how namespaces are handled in the infoset. ... 17/6: No further progress 25/6: Mike has added bit order capability as per action 233. ... 23/9: No further progress
258	**Consider allowing more flexible escapeCharacter schemes (Steve)** 6/5: Motivated by example of an escape character which is active when in front of an in-scope delimiter, but not when in front of another character. 20/5: Can't model Mike's example with current facilities, but Mike's example is a generalisation of a particular MITRE example. Do we really need this? Jonathan to follow up. 3/6: ~~Closed~~. Jonathan has provided the background to the MITRE example which was really about initiators and terminators. The generalised use case is perhaps speculative, so it was agreed not to change the DFDL spec to handle this unless a concrete use case emerges. 17/6: **Re-opened.** vCard 3.0 (http://tools.ietf.org/html/rfc2426) is an example of a format that exhibits the need for this. Need a proposal to handle this case, and which fits in with the existing extraEscapedCharacters and escapeEscapeCharacter property. Noted that using lengthKind 'pattern' is sometimes a way of working round this kind of thing. ... 15/7: No progress 22/7: Steve has started to write up a proposal. ... 26/8: No further progress 2/9: Strawman proposal sent by Steve for comment. Concern over name of new property. Review for next week. ... 16/9: No further progress 23/9:
260	**Positional and non-positional sequences (Steve/Tim)** 10/6: Spec defines the above but also allows different occursCountKinds within the same sequence which may have different (implied) separatorSuppressionPolicy, which results in a sequence which is a mixture of both. Should this be allowed? If so what are the rules? Can certain combinations be disallowed? 17/6: IBM have discussed internally and will submit a proposal. 25/6: Proposal sent to WG. Initial reaction is that the intended semantic for Positional sequence is option a) - an observer of the raw data can identify an occurrence of an element in the sequence solely by counting separators. 8/7: Tim emailed an example which he would like to discuss before WG decides on option a) or

	b).	
	...	
	22/7: No progress	
	29/7: Discussed background, Tim and Steve to have a position for next call.	
	4/8: IBM has held meeting, Steve to write up findings	
	26/8: Revised proposal sent by Steve - please review. Need to decide if included in next published spec.	
	2/9: Concern from Tim that the proposal still allows 'hybrid' sequences that mix 'positional' and 'non-positional' separator suppression characteristics. The alternative is to prohibit certain combinations of occursCountKind and separatorSuppressionPolicy. If this makes implementation materially easier then it is a possibility, but care must be taken to minimise breakage of existing schemas. Tim to make new proposal. Separately, clarification needed on behaviour of maxOccurs '0' and occursCount '{0}' - see **new action 271.**	
	...	
	16/9: With IBM	
	23/9: IBM holding meeting prior to WG call	
262	**Publish updated specification (All)**	
	10/6: Start to address Word comments in draft r11. Got to the start of section 12.3 (length properties).	
	17/6: Extra call held, nearly all comments addressed in draft r12 by Mike. Draft r13 created by Steve for remaining comments plus recent spec errata. Mike aiming to create draft r14 for WG review by next week.	
	25/6: Mike fixing up the references to ensure hyperlinks present in derived html. Then draft r14 can be issued for WG review.	
	8/7: Mike has posted draft r15 on Redmine. Please review for next call.	
	15/7: Mike and Steve making minor edits, Mike will merge in Steve's. Mike will make the us-ascii-7-bit-packed change noted above in action 233.	
	22/7: Draft r17 posted on Redmine. Decided that it makes sense to include the changes from action 233 and 224 in the current draft, as these were raised as public comments, the work is nearing completion, and there is demand from Daffodil sponsors.	
	29/7: Draft r18 posted on Redmine. IBM to review asap.	
	4/8: Draft r20 posted on Redmine. Mike to create draft r21.	
	26/8: Steve reviewed draft r21 and made editorial changes. Mike to create draft r22 which will include the bitOrder changes from action 233.	
	Discussed what other open actions need to be added before spec can be published. Possibly actions 260, 261, 269.	
	2/9: Draft r22 posted to Redmine. All please review.	
	9/9: Draft r22 reviewed and approved. Mike creating draft r23 to include recent actions. OGF board have been notified to expect spec revision. Mike and Steve to send assembled collection (draft r23, experience docs 1-3, obsolete GFD.174).	
	16/9: Draft r23 created. OGF Steering Group have been sent the collection of documents, will review at their next meeting. Experience 3 (bit order) now in public comment. OGF were expecting Redmine Issue trackers to be used to track changes to the spec arising from public comments. This makes the process more visible to reviewers. Steve has retrospectively created Issue trackers for all changes resulting from public comments and feedback since then (http://redmine.ogf.org/projects/dfdl-wg/issues) so that the review is not held up. DFDL WG will raise Issue trackers for all future spec issues that result in a non-editorial change to the spec.	
	23/9:	
265	**Proposed new lengthKind to handle stop-bit encodings (All)**	
	22/7: As used by Fix Adapted for STreaming (FAST) and others.	
	29/7: No progress	
	4/8: Is this best considered as over-punching or use of 7-bit chars?	
	...	
	23/9: No further progress	
267	**Allow lengthUnits 'bits' for more representations (Mike)**	
	29/7: Mike proposes allowing this for all text reps and binary floats and doubles. What about	

	decimal and hexBinary? What about complex elements? ... 23/9: With Mike, low priority.
268	**Missing XPath functions (Mike)** 29/7: XPath string-to-codepoints() and codepoints-to-string(), plus more DFDL specific bitwise functions. ... 23/9: With Mike. Jonathan mentioned the EXPath:Binary module, might be informative.
270	**Investigate ICU handling of embedded whitespace when lax parsing (Richard)** 26/8: Spec says that it is treated as zeros, not sure that is correct. 2/9: Richard @ IBM has found different behaviour from ICU for Java and C. Bug tracker opened http://bugs.icu-project.org/trac/ticket/11230. 9/9: No update to ticket 16/9: Ticket updated. ICU-C behaviour thought to be correct, where the space is treated as ending the number. ICU-J appears to be treating the whitespace as a grouping separator. 23/9: Some activity on ticket, no concrete conclusion yet.
271	**Use of maxOccurs '0' (All)** 2/9: Legal in XSDL and DFDL. One use case in XSDL is when deriving a complex type by restriction, where it is used to indicate that an element in the base type must not appear. Another use case could be if a schema undergoes a version revision that removes elements; perhaps clearer if the removed element gets maxOccurs '0' rather than omitting it. Steve has seen an instance of its use in a DFDL schema. So if DFDL continues to support it, need to document the behaviour for the various occursCountKinds and what happens to separator. Also behaviour of occursCount expression that returns 0. Proposal needed. ... 23/9: No progress

Closed actions

No	Action

Deferred actions

No	Action
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
200	**Establish recommended practices for pushing changes to GitHub (Mike)** 29/1: Mike will talk to Tresys who have used Git a lot. 5/2: Mike to talk to Tresys this week, Tim has sent some links. 12/2: Information sent by Mike, Steve to review. ...

	2/12: No further progress 14/1: Deferring until needed
241	**Public comment : Bi-di properties placement in precedence section (All)** 7/11: This looks deliberate but the asymmetry between parsing and unparsing is unclear . Really needs Daffodil or IBM DFDL to implement these properties, which has not happened yet. Deferring this action.
242	**Public comment : dfdl:valueLength and dfdl :contentLength descriptions (Mike)** 19/11: http://redmine.ogf.org/boards/15/topics/63. Agreed that the function names were ok as per errata 3.18, and that the spec is clear that they refer to the grammar regions. However the grammar regions mentioned do not fully include literal nil values . Discussed what happens when parsing - remember the length or re-parse? What about lengthUnits 'characters' when the data is binary? Also the 'Notes' that follow the table need to be reworked. 26/11: Needs wording to handle all the issues found , assigned to Mike. ... 11/3: Still with Mike 25/3: Mike has sent out revised wording, reviewed by Mark and Steve. Noted that the words need to explain the concept of building a complex element from the bottom up , and these words are equally applicable to several places in section 12.3. Mike to revise accordingly . 11/4: More revised wording sent by Mike. Started to review but realised it needed some off-line preparation and thought. Review for next call. 15/4: Review comments from Steve and Tim. The functions need to be clear that they work off the infoset value. The detailed wording is needed but should be removed to a new sub-section of 12.3, probably at end. Most sub-sections of 12.3, and the functions in 23.5.3 will refer to this new sub-section. 23.5.3 should limit itself to behaviour specific to the functions , such as not potentially represented, the effect of the $lengthUnits argument. Also discussed what happens if $path argument returns a nodeset > 1; should be a processing error, can always use a predicate to select one node of an array. 29/4: See various email discussions . Several things noted by Mike, and he recommends a rewrite of some of section 12.3. Then the description of the two functions becomes much simpler. Deferring for now, and will resurrect after current spec revision is finalised . 6/5: Mike is working on a mind map for the length section
251	**Create official error codes (All)** 5/2: Create official error codes for all possible errors implied by the DFDL spec . This is a big piece of work, so this action is deferred for now . Was formerly part of action 066.

Work items :

No	Item	Owner	Target	Status