DFDL WG Call Minutes

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

Meeting about Meetings\OGF

Project	**DFDL 1.0**
Meeting Date	**04-Aug-14 (Mon)**
Meeting Time	**15:00 - 16:00**

Created by Steve Hanson on 09-Mar-11
Last Modified by Steve Hanson on 04-Aug-14

OGF DFDL Working Group Call , 4 Aug 2014

Agenda

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

1. Daffodil Open Source Project
 Status update.

2. 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
Mike Beckerle
Mark Frost
Andy Edwards

Apologies
Tim Kimber

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

Minutes

1. Daffodil Open Source Project
Release 14 under test, incorporating performance improvements and bitOrder proof of concept.

Meeting closed
16:05 UK

Next regular call
Due to vacation, Tues 30th Aug @ 16:00 UK, unless something crops up.

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: **270**

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.

...
4/8: No further progress

233	Public comment : Formats with bit order reversed (Mike)

Public comment : Formats with bit order reversed   (Mike)
1/10: http://redmine.ogf.org/boards/15/topics/43. Mike to provide words for potential new
property for review.
8/10: Words sent by Mike generated considerable discussion. Mike will update the words to
make the subject more consumable, and move the bulk of the discussion to a new main section
at the end of the spec (suggest between existing sections 24 & 25).
22/10: Mike wants to have a working implementation before closing on this, so marking the
public comment as deferred.
31/10: Deferring for now
25/6: Un-deferring as Daffodil has implemented this now and Mike has created an experience
document in Redmine (). All to review Mike's document before next call on 8th July, and
respond with comments. IBM will hold a review and send consolidated comments.
8/7: IBM has reviewed the experience document and sent back comments on the 'required
changes' section. Discussed the comments. It was suggested that the 7-bit encoding enum and
the new byte order enum were made x- values and therefore implementation defined. Clarified
that new bitOrder property applies to same binary reps as byteOrder and not to text reps as a
different bit order for an encoding is considered a different encoding. Also wondered whether
bitOrder property could be incorporated into a x- byteOrder enum, if the concept is not
considered to be orthogonal, as the addition of new scoped properties causes existing schemas
to become technically invalid. Mike to weigh up pros and cons of all these.
15/7:
i) Mike would prefer that bitOrder and new byteOrder enum are part of the base DFDL spec.
IBM is ok with this.
ii) Discussed best way to incorporate the us-ascii-7-bit-packed encoding. Agreed that a
separate OGF DFDL document should document any encoding names that are not IANA
names or CCSIDs, rather than the spec itself. Also agreed that all such names should start
'x-dfdl' to avoid clash with IANA names. **Errata 2.95 and 2.107 to be updated to reflect this** .
Change to be made to spec draft r15 to remove specific reference to us-ascii-7-bit-packed.
iii) IBM suggested an alternative to proposed lengthKind 'fixedOrTerminated' where the simple
element is wrapped in just a sequence. Mike also has an alternative involving a hidden group
and optional elements. Mike to consider preferred solution.
iv) Mike to respin his experience document. IBM will continue its review with that.
22/7: Revised MIL-STD-2045 experience document posted on Redmine. It includes an excerpt
from the MIL-STD-2045 spec which shows what is meant by LSBF. It is not reversal of bits on

the wire. It means groups of bits are packed into bytes starting from the LSB (ie, what is typically viewed as the right-hand end). Within a group of bits the order of the bits is same as normal. IBM sent a further example which hopefully illustrates this more clearly. Mike will split his document into two - bitOrder, byteOrder, encoding as the first, the other proposals as the second. The bitOrder and byteOrder changes will also be added to the current spec draft.
New document for non-IANA/CCSID encodings posted to Redmine. Reviewed by Steve and Andy, comments sent. For purposes of current spec draft, the content will be included in a spec appendix (long term a separate document will be needed).
29/7: Documents have been restructured as per Mike's email. IBM is concerned that the WG's understanding of the DFDL changes necessary to support MIL-STD-2045 keeps changing. Mike has published an email from an SME to address this concern. He is confident that DFDL with bitOrder and 7-bit/6-bit encoding support can handle the majority of these formats. Link-16 spec is 8000 pages so not possible to guarantee that there are not minor issues. IBM to review the reworked document that only discusses bitOrder. Noted that the combination of bigEndian & LSBF has not been seen so far.
IBM has reviewed and commented on the MIL-STD-2045 additional features. This will be discussed under **new action 266.**
Tresys has published a draft set of schemas on DFDLSchemas GitHub for MIL-STD-2045. Mike will update the schemas to reflect IBM's comments.
4/8: IBM has reviewed Mike's reworked document. It is becoming clear that bitOrder is not just applicable to binary content, it applies to all parts of the grammar, and so at a very low level. Implies that bitOrder is an explicit property on all objects, including text reps (and hence should not be in the DFDL standard encoding template). Need to ensure that DFDL's compositional properties are not compromised. Suggested that switching bitOrder automatically aligns to new byte, rather than gives an error. Mike to revise document. Noted that the 7-bit and 6-bit ASCII DFDL standard encodings should work in both MSBF and LSBF.

248	Discriminators and potential points of uncertainty (Steve)

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.

	... 4/8: No 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. ... 4/8: No further progress
258	Consider allowing more flexible escapeCharacter schemes (Mike) 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. ... 4/8: No further progress
260	Positional and non-positional sequences (All) 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
261	Implied separatorSuppressionPolicy for occursCountKind 'expression' (All) 10/6: Spec says it is 'never' (positional sequence) but you have to parse to identify the position,

	so isn't that non-positional? 17/6: Some other issues noted around 'expression' as per email thread. IBM have discussed this internally and will submit a proposal. 25/6: Proposal sent to WG. The aim of ensuring the integrity of the data stream is sound, but need to think carefully about the use of dfdl:contentLength() in outputValueCalc to make sure there is no deadlock or inconsistency. 8/7: No progress 15/7: Alex concerned about the dfdl:contentLength() and outputValueCalc scenario. He will respond to the email. ... 29/7: Still with Alex. Noted that same principle applies to lengthKind 'pattern' but problem with look-ahead means that applying the regex on unparsing will not always work. 4/8: IBM has held meeting, Steve to write up findings
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.
263	DFDL WG @ OGF 42 (Steve/Mike) 15/7: Want to demo both DFDL implementations. 22/7: No further progress 29/7: IBM has confirmed travel approval for Steve to attend. Agreed to host a presentation followed by a tutorial, which will use IBM DFDL editor to create schemas for a couple of formats that can then be parsed by both IBM DFDL and Daffodil. PCAP and vCard suggested. Attendees can pre-download the material and follow-my-leader if desired. Steve to confirm with Alan Sill. 4/8: Alan Sill ok with format. Daffodil team to create kit to enable easy demonstrating of Daffodil.
264	Use of ES entity in delimiters (Steve L) 15/7: Proposal to allows ES in delimiters subject to same restrictions as WSP* as inconsistent otherwise. Due diligence needed to see whether there is anywhere else in the spec that could also allow ES because it also allows WSP*. 22/7: With Tresys 29/7: Still with Tresys. Noted that the IBM's comments on the MIL-STD-2045 additional features document used this feature to negate the need for lengthKind 'fixedOrTerminated' (though WSP* would also work). 4/8: Still with Steve L
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?
266	Additional features useful for modelling MIL -STD-2045 (All) 29/7: Now a separate document on Redmine. IBM has reviewed the document, and suggested ways to model some of the more awkward constructs using existing facilities, but there are still a couple of points to discuss.

No	Action
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? 4/8: 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. 4/8: With Mike. Jonathan mentioned the EXPath:Binary module, might be informative.
269	**dfdl:occursIndex() function (Steve)** 29/7: Clarify behaviour and decide whether an argument is needed to make it context independent. Noted that fn:position() also does not take an argument, but it returns position in current sequence and not position in array. 4/8: With Steve but consensus is that an argument is needed.

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
045	Resolve public comments and incorporate into spec GFD.207	All	2014-04-30	Pending