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	07-Feb-17 (Tues)
Meeting Time	15:00 - 16:00

Created by Steve Hanson on 09-Mar-11
Last Modified by Steve Hanson on 07-Feb-17

OGF DFDL Working Group Call , 7 February 2017

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. NITF image file schemas now available via https://github.com/DFDLSchemas

2. dfdl:lengthUnits for hexBinary
Daffodil team suggest alllowing lengthUnits 'bits' for hexBinary data to support use cases.

3. Unparsing choices within hidden groups
Daffodil team believe this is problematic as the infoset can not be used to resolve the choice .

4. DFDL schema for Praat TextGrid
Authored by Kristian Kankainen, candidate for https://github.com/DFDLSchemas

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
Steve Hanson
Mike Beckerle

Apologies

Minutes

1. Daffodil Open Source Project
NITF image file schemas now available via https://github.com/DFDLSchemas.

One of the image formats is jpeg which causes alignment issues for Daffodil. Being worked.
Schemas for several image formats in production via MITRE.
TIFF is a problem as it uses offsets internally.
Mike presented on DFDL and its applications at Cross-Domain Technology Forum.
More interest now that Daffodil has its unparser.

2. dfdl:lengthUnits for hexBinary
Daffodil team suggest allowing lengthUnits 'bits' for hexBinary data to support use cases.
Can see the reasons, need a fully worked proposal to evaluate. **New action 292.**

3. Unparsing choices within hidden groups
Daffodil team believe this is problematic as the infoset can not be used to resolve the choice.
May require existing restriction on dfdl:outputValueCalc not allowed on head of branch to be removed. May require new properties for unparsing of direct dispatch choice. Needed for Link-16.
New action 293 raised to investigate.

4. DFDL schema for Praat TextGrid
Authored by Kristian Kankainen, candidate for https://github.com/DFDLSchemas.
Kristain invited to next call to describe the format and use cases.
Schemas developed using Daffodil. Steve will test updated schemas with IBM DFDL.

5. IBM HLASM now has ability to generate DFDL schemas for DSECTS
Steve to see if this can be publicised.

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

Meeting closed
16:05 UK

Next regular call
21st February @ 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

Subject	Document Type	Created	Modified

Next action: **294**

Actions raised at this meeting

No	Action
292	**Write up proposal for allowing hexBinary elements to have dfdl :lengthUnits='bits' (Mike)** 7/2/2017: Mike will create a proposal for evaluation.
293	**Investigate solutions to enabling choices in hidden groups to be unparsed (All)** 7/2/2017: Study of problem needed in order to best evaluate any proposals.

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. ... 31/3: No further progress 14/4: Agreed that the need for better tutorials has become pressing for Daffodil users who aren't using IBM's tools and material. Discussed creating tutorials based on a tdml file with comments that is processed to produce html. Mike to investigate. 28/4: Mike has sent an example tdml file which embeds instances of a new 'tutorial' element in various places. These elements contain html which can be extracted and formatted in a browser. Suggest future DFDL tutorials are created using this technology. 12/5: Not discussed ... 22/9: No further progress 3/11: Daffodil team has someone working on the new 'tutorial' element in tdml files. In time this should result in some new tutorials and re-working of existing tutorials. 5/1/16: Mike has started a bitOrder tutorial using the tdml file approach (uses stylesheets to render html). 16/2: The bitOrder tutorial is available on the web @ https://opensource.ncsa.illinois.edu/bamboo/artifact/DFDL-MASTER21/JOB1/build-132/Tutorials/bitorder.tutorial.tdml.xml 1/3: Awaiting review. Web-based try out facility under development at Tresys. ... 7/2/17: No further progress
242	**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. ... 9/12: No further progress 6/1/15: Mike to resurrect this as Tresys would like to run their tdml suite against both Daffodil and IBM DFDL. ... 10/2: No further progress 24/2: Mike updating the Daffodil TDML test runner to handle unparser (ie, serializer) tests

	... 14/4: No further progress 28/4: Tresys have enhanced their tdml runner to allow unparser tests and round-trip tests (parser->unparser->parser) as well as the new tutorial tag (see action 228) 12/5: Not discussed ... 3/11: No progress 5/1/16: No progress. Needs more interoperability between implementations to be really useful. ... 7/2/17: No further progress
250	**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. Deferring until needed. .... 23/9: Rewrite should be postponed to future 1.1. Still need to answer the original questions about the functions though... ... 25/4/2016: Undeferring action as some of these issues are now impacting Daffodil team as they write their unparser. Steve has sent the email threads on this action to Mike. Mike will combine with his issues and distill into a single thread. 7/6: Thread to include use of a variable with dfdl:outputValueCalc. May be undefined at point of evaluation. 5/7: Mike has been looking at this. Two main points: 1) Computing the content length of a complex element with internal alignment. 2) Computing length in chars of a complex element which is not 100% text Mike will send out a discussion via email. Noted that rules should also apply to prefixed length calculation. (Aside: IBM DFDL unparser does not support a prefixed length complex element with length units chars and variable width encoding). 2/8: Mike has sent out several emails. #1: Proposes that term ComplexValue is added to the grammar to better handle ElementUnused. Knock-on effect on the wording of dfdl:valueLength(). Agreed on the change. **Issue https://redmine.ogf.org/issues/316 created.** Spec says escapeCharacter, escapeBlockStart, escapeBlockEnd, escapeEscapeCharacter

	contribute to the content length of an element. This is not correct, they are part of the value length. **Issue https://redmine.ogf.org/issues/317 created.** #2: Gives some options for computing DFDL length functions when target complex element has interior alignment. Agreed that the DFDL processor should detect this and give runtime SDE. This is an example of expression forward reference deadlock. Need new paragraph in section 23 of spec to cover this. **Issue https://redmine.ogf.org/issues/318 created.** #3: Argues that DFDL should only encode/decode when it needs to when computing DFDL length functions, to allow for performance. Agreed that this behaviour was ok. **Issue https://redmine.ogf.org/issues/319 created.** 13/9: Review issues created. Need to come up with the revised descriptions for dfdl:valueLength and dfdl:contentLength functions before action can be closed. ... 7/2/17: No further progress
279	**Improve defaulting description to explicitly cover local groups (Mike)** 28/4: Only talks about elements, should mention local sequence and choice. 12/5: Not discussed 23/6: Section 15.1.3 needs to say what happens when a choice branch does not contain any elements; such a choice branch is selected (but see action 280 below as minOccurs '0' might change this). Section 9.4 also needs updating to say what happens when local groups are found within a complex type. 11/8: Steve did some tests with IBM DFDL. Just need some words as above. Action assigned to Mike. 25/8: In progress ... 5/1/16: No progress ... 7/2/17: No further progress
287	**Find a way to handle a variable path step in DFDL expression (All)** 1/3: DFDL4S currently using a hack that embeds a regex in a path step. 10/5: No progress 24/5: Need example from DFDL4S .. 5/7: Need to ask DFDL4S for example. 2/8: DFDL4S sent example. They use dfdl:contentLength() with a path that has a step that contains a regex as a wildcard. Mike has requested the wider set of schemas to be sent, in order to see if there is a viable alternative. 13/9: Response received from DFDL4S, not yet analysed. ... 10/1/17: No further progress 7/2/17: Mike has analysed the schemas and sent a comprehensive reply to DFDL4S. He believes that the variable path step is effectively a way of parameterizing the expression, and has described how this can be done using DFDL variables. DFDL4S have responded and will talk to the contractor that authored the implementation.
288	**Decide on error function specification (Mike)** 5/7: An error function is useful for expressions when unparsing. Mike to evaluate XPath fn:error(). 2/8: Mike looked at fn:error(). Concern about use of QName argument, but as QNames are used elsewhere in DFDL this should be ok. Agreed that fn:error() looks suitable. 13/9: Presumably fn:error() throws a processing error. Do we need to be able to throw a schema definition error too? Mike to think about it. Noted that QNames using a prefix maps to the DFDL namespace should be reserved. 11/10: Prime use case is for unparsing which does not need to distinguish error types 8/11: Mike to decide whether both error types needed, and to write up 10/1/17: Noted that Daffodil has implemented this in the meantime. **Issue https://redmine.ogf.org/issues/324 created**. Issue needs resolving and documenting. 7/2/17: Daffodil needs to update their implementation as today it takes no parameters.

289	**Unparsing: expression refers backwards to outputValueCalc which refers beyond it .** 2/8: Need to decide if this is allowed and if so if there are any restrictions. 13/9: Motivating scenario is where a variable is being set to a length element using dfdl:setVariable, which on unparse is set using dfdl:outputValueCalc. So although the variable is referring backwards to the length element, it is effectively forward referencing so must block. Mike believes this is unavoidable. 11/10: Daffodil has implemented this, Mike to provide scenario 8/11: Mike couldn't find example, will continue to look 10/1/17: Mike has realised that all the examples were reworked to avoid using variables, hence why can't be found. 7/2/17: Daffodil will soon be implementing dfdl:newVariableInstance which will bring this up again.
290	**Should utf16Width be optional ? (Steve)** 13/9: Adds complexity to implementing the core set of encodings. Steve to investigate IBM DFDL's support for utf16Width to assess the complexity. 11/10: No further progress 8/11: IBM DFDL uses the property to set the min/max bytes per char properties of its internal charset class. In that sense it is no different from any other variable-width encoding. Steve believes that there is no additional complexity beyond that needed for UTF-8. Mike to think further, particularly for the case when surrogate pairs are involved. 10/1/17: Mike believes there are issues. Steve to do some testing with IBM DFDL to see if the implementation works. 7/2/17: No progress
291	**Should complex fixed length element with variable width encoding be optional ? (Steve)** IBM DFDL supports this but with restrictions on the child elements. Where did these restrictions come from? 11/10: No further progress 8/11: Steve believes that IBM added these restrictions unilaterally as the spec wasn't clear on the behaviour. The check is not difficult to implement; just need to set a flag when such an element is encountered, and any children (recursively) check the flag and ensure they comply with the parent (text rep, character units, encoding). 10/1/17: Need to clarify the DFDL spec and decide whether the IBM restriction is valid. 7/2/17: No progress

Closed actions

No	Action

Deferred actions

No	Action
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. ... 23/9: Candidate to be moved out to 1.1 ?

Work items:

No	Item	Owner	Target	Status