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	**10-May-16 (Tues)**
Meeting Time	**15:00 - 16:00**

Created by Steve Hanson on 09-Mar-11
Last Modified by Steve Hanson on 10-May-16

OGF DFDL Working Group Call , 10 May 2016

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. Simple element zero length matching for nil
Section 9.3.2.1 says
 If the result is length zero as described above, the representation is then established by checking, in order for:
 1. nil representation (if %ES; is a literal nil value).
 2. empty representation.
 3. normal representation (xs:string or xs:hexBinary only)
 4. absent representation (if none of the prior representations apply).
But bullet 1 should be:
 1. nil representation (if either %ES; or %WSP*; on its own is a literal nil value).

3. Example of byteOrder 'bigEndian' and bitOrder 'leastSignificantBitFirst '
Mike has encountered a format called Asterix that uses this combination .

4. DFDL interoperability issues
A project in the US is parsing data using Daffodil , then serializing with IBM DFDL.
They observe:
```
  a) Both implementations use alternative methods to implement certain
     functions that are not spec-compliant. DFDL files are not compatible
     between the implementations
  b)) Daffodil output isn't serializable by IBM DFDL
```
Interoperability is key to progressing to REC status so these need to be understood .

5. Date/time functions
IBM has had a customer ask whether the DFDL serializer can insert the current date /time into
an output field. This would require the addition of the XPath functions fn:current-date-time(),
fn:current-date() and fn:current-time() to the DFDL XPath 2.0 subset, which could then be
used in a dfdl:outputValueCalc expression.

6. AOB

<div align="center">

Meeting Minutes
</div>

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
Continuing on Daffodil unparser, including dfdl:outputValueCalc.

2. Simple element zero length matching for nil
Section 9.3.2.1 says
If the result is length zero as described above, the representation is then established by checking, in order for:
 1. nil representation (if %ES; is a literal nil value).
But bullet 1 should be:
 1. nil representation (if either %ES; or %WSP*; on its own is a literal nil value).
Agreed to make the proposed update.
Issue https://redmine.ogf.org/issues/309 created.

3. Example of byteOrder 'bigEndian' and bitOrder 'leastSignificantBitFirst' ?
Mike has encountered a format called Asterix that numbers bits in bytes in the 'reverse' order used by 'LSBF'. However if the bits don't ever span bytes then a DFDL schema can be created which only uses 'MSBF'.

4. DFDL interoperability issues
A project in the US is parsing data using Daffodil, then serializing with IBM DFDL.
They observe some interoperability issues.
Interoperability is key to progressing to REC status so these need to be understood.
Information initially to be sent to Tresys due to confidential nature of data.
Namespaces are a known area where there are differences.
New action 288 raised.

5. Date/time functions
IBM has had a customer ask whether the DFDL serializer can insert the current date/time into an output field. This would require the addition of the XPath functions fn:current-date-time(), fn:current-date() and fn:current-time() to the DFDL XPath 2.0 subset, which could then be used in a dfdl:outputValueCalc expression. Agreed to add the functions.
Issue https://redmine.ogf.org/issues/311 created.

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

Meeting closed
16:40 UK (after delayed start)

Next regular call

Tues 24th May 2016 @ 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: **288**

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.
...
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.
10/5: No further progress.

250	Standardise on a single tdml format for DFDL tests (All)

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. ... 10/5: No 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 ... 10/5: No progress
282	**Does XPath have operators for checking a value is in a range ? (Steve)** 12/5: Investigate whether equivalent to DFDL4S 'in' operator exists. 23/6: Mike has found an XPath 'intersect' operator. Handles the enumeration case well, but not as convenient for ranges as DFDL4S's 'in' operator. 11/8: Looked back at the motivating example from DFDL4S. Agreed that DFDL functions to do the equivalent of 'in' and 'inrange' would be useful if nothing can be re-used from XPath. Steve to write up a proposal taking into account different data types. ... 22/9: Proposal sent by Steve for new functions dfdl:checkValues(), dfdl:checkRangeInclusive(), dfdl:checkRangeExclusive(). Discussed whether both range functions needed, and whether they should be allowed for float and double. Mike noted that dfdl:checkConstraints() and simple types could be used instead of all three functions if values were static. Follow-up with DFDL4S team to see if they had thought of that. 3/11: DFDL4S not happy with usability of dfdl:checkConstraints type, but ok with the intersection operator. Next step is to see what the DFDL4S schema would look like if rewritten to use dfdl:checkConstraints and intersection. 5/1/16: Agreed that dfdl:checkConstraints() is not ideal as it requires creation of union simple types. Steve to go back to his proposal from 22/9/15 and rework to include use of the intersect operator. 16/2: No progress 1/3: Discussed with ESA team. Steve to rework his proposal from 22/9/15. 10/5: Reworked proposal mailed to WG. DFDL4S to review for next call please.
286	**Create table that compares Daffodil features to IBM DFDL features (Steve/Mike)** 16/2: Take table from section 22 of DFDL 1.0 spec and add columns showing implementations. 1/3: No progress 10/5: No progress. Work on this for next call.

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

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 ?
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 . Deferring until needed. .... 23/9: Rewrite should be postponed to future 1.1. Still need to answer the original questions about the functions though...

Work items :

No	Item	Owner	Target	Status