## DFDL WG Call Minutes

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

| Project | DFDL 1.0 |
| ---: | :--- |
| Meeting Date | 12-Nov-20 (Thurs) |
| Meeting Time | 16:00-17:00 |

Created by Steve Hanson on 09-Mar-11
Last Modified by Steve Hanson on 12-Nov-20
OGF DFDL Working Group Call, 12th November 2020

## Agenda

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

## 1. Implementations Update

2. $A O B$

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
Bradd Kadlecik
Mike Beckerle

## Apologies

## Minutes

## 1. Implementations Update

Not discussed

## IPR Statement

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

## Meeting closed

17:10 UK

## Next regular call

10th December 2020 @ 16:00 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

## Actions raised at this meeting

| No | Action |
| :--- | :--- |
|  |  |

## Current Actions:

| No | Action |
| :---: | :--- |
| $\mathbf{2 8 9}$ | Unparsing: expression refers backwards to outputValueCalc which refers <br> beyond it. <br> 2/8/16: Need to decide if this is allowed and if so if there are any restrictions. <br> $13 / 9:$ Motivating scenario is where a variable is being set to a length element using <br> dfdl:setVariable, which on unparse is set using dfdl:outputValueCalc. So although the <br> variable is referring backwards to the length element, it is effectively forward <br> referencing so must block. Mike believes this is unavoidable. <br> 11/10: Daffodil has implemented this, Mike to provide scenario. <br> $8 / 11:$ Mike couldn't find example, will continue to look <br> 10/1/17: Mike has realised that all the examples were reworked to avoid using <br> variables, hence why can't be found. <br> $7 / 2:$ Daffodil will soon be implementing dfdl:newVariablelnstance which will bring this |

up again.
17/4/18: Waiting for Daffodil to implement dfdl:newVariableInstance
15/5: Daffodil team have supplied an example of this from the PCAP schema. Likely to require a flag on newVariablelnstance (or maybe variable declaration) to indicate whether needed on parse, unparse or both.

17/10/19: No further progress
12/12: Daffodil starting to implement dfdl:newVariableInstance
...
16/4/20: No further progress
30/4: Daffodil has implemented newVariableInstance so now in a position to look at this again.
9/7: No update
23/7: Daffodil has implemented newVariableInstance so Mike will look at this again
3/9: Daffodil team looking at this and related issues surrounding expression evaluation when unparsing.

29/10: No further progress
12/11: Mike has a scenario in the PCAP schema that uses several outputValueCalc. To simplify the expressions, he would like to use a variable with a value set by newVariableInstance. The problem is that the rules for variables are different from those for outputValueCalc, specifically variables can not refer forwards (because they are evaluated on both parse and unparse) whereas that is the point of outputValueCalc. To ease this would require a change to the spec. WG agreed this is not necessary for DFDL 1.0 but a candidate for 2.0 .
309 Create example scenarios to illustrate offset \& pointer requirements (Bradd) 5/4/19: Daffodil have a draft proposal for offset support, TPF have experimental implementation for pointer support. Need examples to show the requirement, especially unparsing.
2/5: Bradd supplied an example of pointers. On parsing the pointer is used as an absolute address to a piece of accessible memory, and the element is parsed from that location. On unparsing memory is allocated and unparsing of the element occurs into that location and the pointer set to the location (memory allocation is implementation-defined). Note the pointer value does *not* appear in the infoset. Looks like a useful and workable addition to DFDL. Could solve the parsing requirements for TIFF image files. Bradd also has extension for offset, which is like pointer but uses relative location instead of absolute. Both are examples of indirection. A further example could be specifying a file to read. Contrast this with what DFDL has used the term 'offset' for in the past, namely as an alternative property to alignment/skip which allows the parser/unparser to jump directly to a point in the current buffer. These are orthogonal concepts. Noted that parsing of ZIP files may need both. Secure implementations may need to disallow use of pointers and/or offsets unless they can guarantee to fill everywhere with the fill byte. Implementations should also be deterministic. Agreed that recursion not needed to implement this. Bradd mentioned a further concept 'overflows', an example being an array unparsed into a linked list. Pointers proposal needs to be written up as an experimental feature. 31/5: Bradd to write up pointers proposal as an experimental feature.

11/7: No update
8/8: Bradd aiming to get this written up for next time. Also needs issue tracker raising. 29/8: Bradd unable to make the call.
17/10: Written up for review and sent to WG but not as an experimental feature document. Mike also noted http://www.binarydom.com/sdk/doc/bddl.shtml. Mike has reviewed and commented on the write-up, Steve needs to do the same, then send back to Bradd. Main discussion was around unparsing, eg, buffering implications, whether to try and format exactly or canonically. In parallel, Bradd to create an experimental feature document (see table below).

12/12: Bradd sent an updated document. WG will review for next meeting. There was some discussion about the use in the document of the term 'empty' and whether that really meant 'missing'. This led to an in-depth discussion about the different use cases for default values, it is likely that DFDL 2.0 will introduce support for some of these, specifically:

- Item exists in Infoset with default value, so unparse empty rep (the mirror of parsing, as practised by GPB)
- Item exists in data with default value, so remove from Infoset post-validation (the mirror of unparsing, and a requirement from z/TPF who have a post-parse option to do this)
9/1/20: No progress, still needs reviewing.
16/4: Steve \& Mike to review latest document dated 2019-12-12 for next call.
30/4: Spent some time discussing Steve's review comments. Conclusion is that the feature is useful and a serious candidate for DFDL 2.0. The properties seem to be the minimum needed to handle the concepts and known use cases. As this is currently an experimental feature we don't have to get it $100 \%$ precise now, and can impose restrictions that z/TPF users would be ok with (for example, no initiators or terminators allowed; binary indirection types only). Important though that the properties and their application is driven by the grammar, so next step is for Bradd to see how the grammar is affected. It would be nice if all the behaviour could be handled at the same point in the grammar as 'prefixLength' but that might not be possible. Property name 'indirectionEmptyValue' probably needs a better name, eg 'indirectionUnusedRep'.... other suggestions welcome. Or perhaps the dfdl:fillByte of the indirectionType could be used?
9/7: Bradd has sent an updated document which includes a grammar section. The grammar changes need reviewing by Steve. Property 'indirectionEmptyValue' changed to 'indirectMissingValue', but missing has other connotations. Mike not keen on using fillByte.
23/7: Steve reworked the grammar changes, introducing the concept of a repeating IndirectElement as a sibling of DocumentElement, being the target of an indirection, and using 'indirectionUnusedValue' to dis-ambiguate from nil/empty/missing. Steve explained the concept, Mike and Bradd to review with that in mind. Agreed that the indirection type could have prefixed length.
3/9: Agreed that IndirectElement is useful. Mike tried out the proposed properties to model TIFF format but ran into problems with recursion, as the indirection is really a linked list. Not possible to treat the recursive list as an array. Would also require 'backward' offsets. z/TPF have not implemented offsets but think only 'forward' offsets needed for their use cases. TIFF also needs property indirectionBase to be absolute, which is something that can be added to proposal. Bradd to update. (Also discussed 'layering' which is a related subject but was deferred a long time ago).
17/9: Bradd has sent updated document. Need to use correct terminology for 'underlying elements' and 'prior element', and update grammar. Once that done, the proposal looks to be in good shape. Issue tracker https://redmine.ogf.org/issues/363 raised and noted in table below. Next step is for Bradd to create an official OGF experience document.
As TIFF could not be modelled, should see if offsets can be used to model zip files, and if possible find other offset examples.

29/10: Waiting for Bradd to create experience document.
12/11: Bradd has updated the document to define "underlying elements" and "prior element", and updated the grammar to add indirection as per Steve's proposal. Final step is to create an official OGF experience document.
Move DFDL 1.0 spec to Grid Recommendation as per GFD. 152 (AII) 27/6/19: Steve to create experience document for IBM DFDL usage. Mike to do same for Daffodil. Steve to reach out to the DFDL4S team at ESA for them to do the same. Mike to send email to OGF to inform them of the WG's intent and to establish who to deal with going forward.
11/7: Mike has sent email to Steve for review.

8/8: Mike emailed OGF and Alan Sill has replied. He has forwarded the email to the Grid Forum Steering Committee with a request for responses and asked Jens Jensen as VP for Standards to help the WG work through the process. Agreed that a draft should be created that fixes all the known typographical bugs (as noted in tracker 233) - Mike will do this. Mike will follow-up with Jens Jensen also.

29/8: DFDL WG needs to find an External Reviewer as there is no Data Area chair currently. Needs to be independent from IBM and Tresys. Also, need to establish whether the existing document can be edited or a new document is needed - the latter would entail a 6 month delay. Need to get the current errata experience document updated with the set of trackers that we need in 1.0.
17/10: External reviewer is to be Martin Westhead, who was a founding member of the DFDL WG. Assumption is that we can use the existing spec and add the errata to it, although the number of errata is large. Mike is working through the issue trackers, adding them to the errata document at
https://redmine.ogf.org/dmsf_files/13384?download=, and updating the spec if they can simply be added in without further discussion or approval. Several require approval and Steve has reviewed those. Mike has responded to Steve's comments with further changes, ready for another review cycle. One erratum requiring more thought is 5.39 (issue 299) as it introduces the new concept of a choice branch with zero occurrences. 12/12: Steve has reviewed latest errata, no changes required. Note 5.39 is correct as written, no more discussion needed. Mike has incorporated the errata into the DFDL spec. Steve has reviewed the spec, and Mike has further updated it. Current version at https://redmine.ogf.org/dmsf_files/13601?download=. Martin Westhead thinks he can review by end of year. Mike will create an html version of the current version, marked as a draft, for users for convenience, and will investigate best place to host it.
9/1/20: The html version of the updated spec is available, and Martin Westhead has started to review the spec and is providing comments. He has concerns about the first few sections of the spec, particularly with regard to concepts being mentioned before explanation, so requiring multiple passes to understand. Mike considering the best way to fix this. Martin is much happier with the later sections. Mike has found a few more minor typographical errors and is fixing.
16/4: Martin has reviewed up to page 123 but had to stop due to other commitments. Mike has integrated errata 5.60-5.63 into a draft spec which is in Redmine as gwdrp-dfdl-v1.0.5-r12.docx, with change tracking on. Mike then created another version, accepted changes, and has started to update with some of Martin's suggestions, got as far as page 50. This is likely to result in structural changes to the spec. Mike to contact Martin again.
30/4: Mike has now incorporated up to page 123, so all of Martin's comments to-date. Mike has contacted Martin, waiting to hear back. As part of this work, Mike has prototyped the property precedence section as tables. Reviewed and some minor comments, but agreed it was an improvement over simple bullets. Mike had some questions about the precedence which were answered. At the end of the call, Steve wondered whether an alternative approach might be better - relate properties to grammar constructs. Then everything is driven from the grammar.
9/7: No further reviews from Martin Westhead. Looks like Grid Recommendation will be pushed out to 2021, as 6 month Public Comment needed.
23/7: Agreed to look for additional reviewer to help speed things up.
3/9: Martin Westhead has reviewed the entire spec. No need for an additional reviewer. Mike to carry on with incorporating Martin's comments as a high priority. Roger Costello observed a property/enum index would be useful.
17/9: Mike struggling with using MS-Word for spec. Alan Sill has suggested GitBook and has uploaded the latest spec draft to a new OGF GitHub repo and generated one. At first glance it looks good. Mike to evaluate and see if this is an improvement. Mike has also raised the issue of RFC2119 conformance (use of MUST, MAY, REQUIRED, OPTIONAL, etc) in an email to WG. The spec uses 'must' etc in several ways, and not all are related to what an implementation needs to do. Those that are, and are therefore subject to RFC2119, will be upper-cased, and a section has been


|  | problem (and by extrapolation ../.. and so on). So, the question is actually what does it mean to apply these functions with an argument that indicates the current element is, or is inside, the element being counted. More thought needed. <br> 29/10: No further progress <br> 12/11: Mike has thought further and written up in an email. WG agreed that Mike's option 3 - a path to self or parent causes an SDE - is probably the best way forward. Implementation teams to trawl their tests to see if any are impacted. |
| :---: | :---: |
| 317 | Publish some real-world numbers and key use cases for DFDL (AII) <br> 16/4/20: This would help with the standardisation process. <br> 23/7: No progress <br> 3/9: Steve has identified that there are $\sim 1600$ customers of IBM Integration Bus / IBM App Connect Enterprise, but not clear what percentage use DFDL. <br> 17/9: No further progress. <br> 1/10: Bradd has sent a link to an IBM z/TPF use case - <br> https://community.ibm.com/community/user/ibmz-and-linuxone/blogs/mark-gambino1/ 2020/09/25/opening-and-connecting-ztpf-with-rest. This is also a key use case for IBM IIB/ACE. <br> 12/11: No further progress |
| 319 | Clarify positive discriminator evaluation in nested choice scenario 23/7/20: Section 9.3.3.1 states what happens if a processing errors occurs, should also state what happens if discriminator evaluates to true. <br> 3/9: Daffodil fixed a bug in this area as a result of the discussion. 17/9: Issue tracker raised https://redmine.ogf.org/issues/364. <br> 1/10: Not yet included in errata document or latest draft spec. <br> 29/10: Will be added to spec via a Public Comment <br> 12/11: Mike to raise Public Comment. |
| 320 | Move DFDL WG documents from OGF Redmine to OGF GitHub (AII) <br> 17/9/20: Agreed that it made sense to re-base DFDL WG documents into OGF GitHub. Steve believes Redmine does not support TLS 1.2 and is prone to being hacked (advert-like comments appear sometimes in issue trackers). Would need to move documents, WG call minutes, issue trackers, news items and introductory material. DFDLSchemas should remain as a separate GitHub organization, but linked to. 1/10: No move as yet but some folders created by Mike. Concern that can't order by date, so need to get folder structure and names right. <br> 29/10: Might get some impetus if the Public Comment for the updated spec is moved to GitHub? <br> 12/11: Public Comment is GitHub-based. Mike will move Redmine trackers for DFDL 1.0 to (Public Comment) issues in the Editor repo, and Redmine trackers for DFDL 2.0 to issues in the DFDL repo. Documents more problematic as noted above. Need to think about call minutes, news items, archived documents. |

## Closed actions

| No | Action |
| :--- | :--- |
|  |  |

## Deferred actions

| No | Action |
| :---: | :--- |
| $\mathbf{2 5 0}$ | Standardise on a single tdmI format for DFDL tests (AII) <br> 5/2/14: Steve has requested permission for IBM to view / use the Daffodil tdml files, as a <br> precursor to trying to standardise on a common tdml format. Was formerly part of action |

66. 

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
$\ldots$
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.

25/7/17: No further progress
3/10: No further progress although forthcoming work to add packed/zoned numbers may force https://redmine.ogf.org/issues/339 progress

11/12: Expected to look at this in the next month or so
4/9/18: No further progress
16/10: Mike has started work on a TDML runner that can drive a pluggable DFDL implementation, in support of interoperability testing, including IBM DFDL.
1/11: Pluggable TDML runner working. On Github at https://github.com/OpenDFDL/ibmDFDLCrossTester. Schema resolution for IBM DFDL achieved using its schema resolver feature and pointing it at Daffodil's resolver. IBM DFDL sample uses mark() on its input stream but IBM believes this is not necessary. 15/11: IBM DFDL and Daffodil have dependencies on different releases of ICU. Forcing changes to the TDML runner to isolate the implementations under test.
29/11: Good progress on the TDML runner, see email from Mike. The ibmCrossTestRig is not part of Daffodil (because it links against IBM DFDL), but is open source Apache License v2, and is currently in review at
https://github.com/OpenDFDL/ibmDFDLCrossTester/pull/1. Steve needs to talk to IBM legal to check this is ok as it currently modifies IBM DFDL sample code.
10/1/19: Daffodil have removed the modified IBM DFDL sample code. Steve to ask whether IBM can donate tests from the existing IBM DFDL test suite.
7/2: If permitted, the tests from IBM can be used to see how the IBM and Daffodil tdmls have diverged.

|  | 5/4: IBM are permitted to send Daffodil some example tdmls. Steve to send some to Mike for next call. <br> ... <br> 31/5: No progress <br> 27/6: Steve has sent example TDML files to Mike, under the existing IBM Grant of Copyright agreement with Apache. <br> 11/7: Daffodil to adapt their test runner so that it works with the IBM TDML file variant. Action can be closed when this is done. <br> 29/8: No further progress <br> 17/10: Noted that this action had its number accidentally swapped with 242 in October <br> 2016. Renumbered back to 250. <br> 23/7/20: No further progress <br> 3/9: Nothing likely to happen for this action in medium term, so moving to deferred actions |
| :---: | :---: |
| 316 | Proposed new experimental feature 'User defined functions' (Mike) <br> 9/1/20: Version 2.5.0. of Daffodil will include this, needs writing up. Does XPath 2.0 provide for this? <br> 16/4: Nothing in core XPath 2.0 that allows user-defined functions, that is left to XSL etc. A write-up is here: <br> https://cwiki.apache.org/confluence/display/DAFFODIL/Proposal\%3A+Feature+to+Supp ort+User+Defined+Functions. Mike has an example where values need to be normalised for later comparison. Bradd has an example where floating point precision needs adjustment. Steve is concerned that DFDL is straying too far into the transformation layer, maybe "user-defined-functions-for-DFDL" should be a separate spec? <br> 3/9: No further progress. <br> 17/9: Nothing likely to happen for this action in medium term, so moving to deferred actions. |

## Work items:

| No | Item | Owner | Target | Status |
| :--- | :--- | :--- | :--- | :--- |
|  |  |  |  |  |

## Experimental Features:

| No | Title | Origina <br> Action | Issue Tracker | Implem <br> entor | Experience Document |
| :---: | :--- | :--- | :--- | :--- | :--- |
| 1 | Experimental feature <br> mechanism | 301 | https://redmine. <br> ogf.org/issues/3 <br> 40 | Daffodil | https://redmine.ogf.org/dmsf_ <br> files/13594?download= |
| 2 | Elements of type <br> 'xs:hexBinary' with <br> lengthUnits 'bits' | 292 | https://redmine. <br> ogf.org/issues/3 <br> 44 | Daffodil |  |
| 3 | Variable Path Step in <br> DFDL expression | 287 | https://redmine. <br> ogf.org/issues/3 <br> 53 | DFDL4S |  |
| 4 | BLOBs as URIs in infoset | 312 | https://redmine. <br> ogf.org/issues/3 <br> 56 | Daffodil | In progress |
| 5 | Data streaming layers | 304 | https://redmine. <br> ogf.org/issues/3 <br> 41 | Daffodil |  |
| 6 | Interoperability | 307 | N/A | Daffodil | https://redmine.ogf.org/dmsf_ |


|  |  |  |  |  | \|files/13589?download= |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 7 | Empty element parse policy | 306 | https://redmine. ogf.org/issues/3 55 | Daffodil | https://redmine.ogf.org/dmsf files/13596?download= |
| 8 | Integer enums as strings in infoset | 294 | https://redmine. ogf.org/issues/3 54 | Daffodil |  |
| 9 | Bi-directional support (moved out of DFDL 1.0) | 241 | https://redmine. ogf.org/issues/3 57 | TBD |  |
| 10 | Indirection using pointers | 309 | https://redmine. ogf.org/issues/3 63 | IBM DFDL (z/TPF) | In progress |
| 11 | User-defined functions | 316 | TBC | Daffodil |  |

