# 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 | **28-Jan-14 (Tues)** |
| Meeting Time | **16:00 - 17:00** |

Created by Steve Hanson on 09-Mar-11
Last Modified by Steve Hanson on 29-Jan-14

**OGF DFDL Working Group Call , 28 January 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. Next steps for public comments**
Who does what next to get the comments into the next draft of the DFDL spec.
The single MS Word spec document needs to be refactored into a MS Word multi-part document (Mike).
The experience documents that listed the errata need updating to include the comments (Steve).
Actions to be raised to track.

**3. Discriminators and potential points of uncertainty**
See email from Steve. A scenario with discriminators and arrays with minOccurs > 0 means that a DFDL schema can have unexpected and fragile behaviour.

**4. Forward progress requirement**
See email from Tim. Several questions about the intent behind this check and its effect in practice.

**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
Tim Kimber
Alex Wood
Mike Beckerle

Mark Frost

**Apologies**
Jonathan Cranford

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

**Minutes**

**1. Daffodil Open Source Project**
Not discussed.

**2. Next steps for public comments**
The single MS Word spec document needs to be refactored into a MS Word multi-part document. The experience documents that listed the errata need updating to include the comments. New **actions 246** and **247** raised respectively.

**3. Discriminators and potential points of uncertainty**
Discussed the behaviour when a discriminator is on an array with minOccurs = n (n <> 0) and occursCountKind 'implicit'. The first n times round the array, there is no actual PoU so the discriminator actually acts on enclosing PoUs, if they exist. While it is possible to create DFDL schema that exploit this, it is not intuitive and makes the schema very fragile. Steve proposed a rule whereby the discrimination effect of a discriminator does not 'leak' out beyond the nearest potential PoU, regardless of whether this turns out to be an actual PoU. Needs some further, investigation, taking into account the different types of potential PoU and different occursCountKinds. For example, does it even make sense for a fixed array to be a potential PoU?

Tim noted terminology could be improved - a choice **branch** is a PoU, not the choice itself.

New **action 248** raised for Steve to create a proposal.

**4. Forward progress requirement**
The wording in section 16.6 is all about preventing an infinite loop from occurring, but pre-dates much of the work done as part of action 140. Is throwing a processing error the correct thing to here? Agreed that it was, because if the array is an actual PoU then the error simply terminates the array and allows the parser to continue ('parsed' and 'implicit'), and if the array is not an actual PoU then the error invalidates the array and is propagated upwards ('stopValue'). One error was spotted, namely that for occursCountKind 'implicit' the check should not be made until minOccurs has been passed. Also agreed that while it is possible to author a schema where no forward progress is made for many occurrences and then to suddenly advance, by cunning use of variables, it is better to go with the majority use cases and treat no forward progress in an unbounded array as an error.

Alex raise a related optimisation question for 'implicit', namely what if maxOccurs is not unbounded but no forward progress is made? Should the parser terminate the array early, in the same manner as unbounded? Otherwise the parser might loop for a large number of occurrences, adding nothing to the infoset and not advancing, while potentially exhausting resources, before reaching the end. Concrete example of this seen when modelling X12 functional loops. Agreed that this should be treated like the unbounded case, and section 16.6 updated accordingly. There is one case where the infoset could be updated, and that is if array is nillable with %ES; as nil value, in which case NIL would be added to the infoset for each occurrence. Again it was agreed to go with the majority use cases and to ignore this NIL case.

New **action 249** raised for Tim to revise section 16.6.

**Meeting closed**

16:15 UK

**Next regular call**
Wed Feb 5th 15:00 UK  (note changed time & day)

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: **250**

## Actions raised at this meeting

| No | Action |
|---|---|
| 246 | **Refactor DFDL spec into a multipart MS Word document  (Mike)**<br>28/1: Mike to see whether this solves the fragility and size problems that the spec experiences today. |
| 247 | **Update DFDL errata documents to incorporate public comments  (Steve)**<br>28/1: To keep all spec issues documented in one place, the public comments will be added to the errata documents. |
| 248 | **Discriminators and potential points of uncertainty  (Steve)**<br>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. |
| 249 | **Revise the array forward progress requirement  (Tim)**<br>28/11: Tim to revise section 16.6 to correct errors and include 'implicit' bounded use case. |

## Current Actions :

| No | Action |
|---|---|
| 066 | **Investigate format for defining test cases (All)**<br>25/11:IBM to see if it is possible to publish its test case format.<br>04/12: no update<br>...<br>17/02: IBM is willing in principle to publish the test case format and some of the test cases. May need some time to build a 'compliance suite'?<br>24/03: No progress<br>03/03: Discussions have been taking place on the subset of tests that will be provided.<br>10/03: work is progressing<br>17/03: work is progressing<br>31/03: work is progressing<br>14/04: And XML test case format has been defined and is being tested.<br>21/04. Schema for TDML defined. Need to define how this and the test cases will be made public<br>05/05: Work still progressing<br>12/05: Work still progressing<br>02/06: Work still progressing on technical and legal considerations<br>...<br>25/08: Will chase to allow Daffodil access to test cases. The WG should define how implementation confirm that they 'conform to DFDL v1'<br>01/09: IBM still progressing the legal aspect. Intends to publish 100 or so tests as soon as it can, ahead of a full compliance suite.<br>08/09: IBM still progressing<br>15/09: IBM still progressing, expect tests to be available within a few weeks<br>22/09: IBM still progressing, expect tests to be available within a few weeks<br>29/09:Test cases are being prepared.<br>06/10: Some test cases should be available next week. Steve would like to be able to show the test case information at OGF 30.<br>13/10: Still progressing<br>10/11: Legal issues cleared, IBM in process of collecting 100 example test cases, ideally ones that fit the 'extended conformance' of NCSA Daffodil<br>17/11: Work is progressing on verifying the test cases. It should be possible to distribute to the WG in 2 weeks.<br>24/11: About half the test cases have been completed and are being reviewed internally.<br>01/12: Test cases should be available shortly<br>08/12: The test cases are in internal IBM review. Probably need a bit of reorganising before publication<br>Stephanie gave a brief overview of the format of the test cases.<br>15/12: Ruth joined the call to provide the latest status. The test cases have been updated and a draft read.me produced. Although not ready for public distribution Ruth will send them to Joe for feedback.<br>22/12: Test cases were sent to Joe for initial testing which found some problems in the Daffodil parser<br>12/01: All current tests use a default format whih Daffodil doesn't currently support. Joe suggested that there should be test that defined the same function using different definition forms. Also suggested that default formats should be provided by the WG. This had always been the intention. Action 133 raised to track.<br>19/01: There is currently no resource available in IBM to make more tests available. IBM to discuss how/if it can make a 'minimal compliance test suite' available.<br>26/01: Action kicked off within IBM. There was a brief discussion abot naming and organisation of test cases but no preferences were expressed |

02/02: IBM will not have the resources to develop a full test suite in the near future. Steve suggested that we produce a list of required test cases so that anyone could supply them.
09/02: Steve had previously sent a list of areas to be tested. Please review.
23/02: Please review Steve's list of areas to be tested
02/03: Alan had reviewed Steve's list and we went through his comments. Agreed there is no need for separate tests for the infoset or for dfdl: property lists, unions etc but comment will be added that these should be exercised during property testing.
09/03: Alan updated the test document. Need more introduction and perhaps adopting the OGF template.
30/03. Ownership of test document passed to Steve. This action is merged with 112 and will cover all aspects of compliance suite.
13/04: IBM will not have time to create a compliance suite in the near future. Probably best to make this action deferred for now.
...
10/07/2012: Discussed schemes to create interchangeable tests. Ideally need a DFDL defined error code per failure, in conjunction with specific inserts.
**...**
26/3/2013: Resurrecting deferred action.
We have got to the point where it makes sense to converge the IBM DFDL and Daffodil variations of .tdml file.
Steve to seek permission from IBM to make the list of IBM DFDL error messages available to DFDL WG.
...
24/5: No further progress.
28/5: Mike summarised the status of Daffodil's tdml runner. Since IBM shared the tdml format, Daffodil has added a) bit file support with in-line comments; b) embedded schema; c) failure checking by multiple string matching. IBM has added a) some flags that map to parser API 'features' such as optional checks; b) code to handle illegal XML characters. 1200 parser test cases written for Daffodil, about 60 of the original IBM shared tests now pass in Daffodil. Steve will email OGF and ask if there is an approved process for demonstrating that multiple implementations generate the same set of test results. To progress with a shared tdml format, IBM will need to get legal approval to view the Daffodil source test cases, Steve to kick this off. Mark noted that IBM's tdml format has evolved in order to make the infoset comparison easier, Mark will see whether the shared tests use the latest version.
4/6: Steve has emailed OGF for guidance, reply received. Experience documents needed to verify conformance, but there is not a requirement to have executable tests. However, a set of executable tests is what we need ideally.
Discussed error messages and identifiers for different errors and what the granularity should be. Steve has asked for permission to send the IBM DFDL error messages to the DFDL WG, they should be used as a starting point. Need to agree what constitutes the minimum content of an error message.
...
22/10: No further progress
31/10: Steve has permission to send IBM DFDL's error messages to the WG.
5/11: Error messages sent
...
2/12: No further progress
14/1: Steve had started to pursue approval for IBM to be sent the Daffodil source test cases. He will resurrect this.
21/1: Noted .tdml format divergence between IBM and Daffodil, best example being error message specification.
Steve explained the IBM DFDL error message prefix CTDxnnny, where x is component (can be ignored), nnnn is unique message number, and is E(rror), W(arning) or I(nformational). If Daffodil used the same message numbers, a .tdml runner could work with both IBM DFDL and Daffodil by looking for the message number in (say) the first 10 chars.
28/1: IBM error messages now in Daffodil source control for future use.

| | |
|---|---|
| 224 | **Add section for implementation defined limits (Jonathan)**<br>3/9: Several places in the spec cite this, should be grouped. Currently partially listed in section 2.6.<br>Also note distinction between 'implementation defined' and 'implementation dependent'. Check spec for correct usage.<br>Resolve during public comment.<br>10/9: No progress<br>17/9: Jonathan sent a reference to the W3C XProc standard where the distinction is made clear. Jonathan will go through the spec and gather everything that is implementation defined/dependent. Public comment to be raised<br>24/9: With Jonathan to raise.<br>1/10: Public comment 97 raised (http://redmine.ogf.org/boards/15/topics/97)<br>8/10: With Jonathan to provide words.<br>22/10: Jonathan has defined implementation defined/dependent and started to classify. Steve and Mike had trouble with the definitions, Steve to re-word and send for comment.<br>31/10: Reworded version sent<br>5/11: Rewording approved. Jonathan proceeding with classification, will distribute for review when complete.<br>...<br>28/1: Still with Jonathan |
| 228 | **Review set of tutorial lessons (All)**<br>17/9: Lesson 1 proposes a set of lessons, needs reviewing as over 2 years old.<br>...<br>22/10: No progress<br>31/10: Becoming a focus for Tresys. Steve to send his 'Modeling Data Formats using DFDL' powerpoint.<br>...<br>19/11: No further progress<br>26/11: Possibility of help from MITRE high-school student, and from Marisa at IBM.<br>...<br>28/1: No further progress |
| 242 | **Public comment: dfdl:valueLength and dfdl :contentLength descriptions (Mike)**<br>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.<br>26/11: Needs wording to handle all the issues found, assigned to Mike.<br>...<br>28/1: Still with Mike |

## Closed actions

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

## Deferred actions

| No | Action |
|---|---|
| 131 | **Transformation of DFDL properties to a canonical form (Joe)**<br>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.<br>15/12: Alan tested against test dfdl schema which worked correctly (after fixing some errors in the schema) |

| | |
|---|---|
| | 22/12: no update<br>12/01: Joe has some defects to fix before making available on gridforge.<br>19/01: There is a difficult problem to solve before Joe make the style sheet public<br>26/01: Working on problems<br>02/02: no progress<br>09/02: As it wasn't a simple as exoected this will be treated as a low priority action<br>23/02: Low prioity<br>09/03: Low priority<br>30/03: Deferring for now |
| **200** | **Establish recommended practices for pushing changes to GitHub (Mike)**<br>29/1: Mike will talk to Tresys who have used Git a lot.<br>5/2: Mike to talk to Tresys this week, Tim has sent some links.<br>12/2: Information sent by Mike, Steve to review.<br>...<br>2/12: No further progress<br>14/1: Deferring until needed |
| **233** | **Public comment: Formats with bit order reversed (Mike)**<br>1/10: http://redmine.ogf.org/boards/15/topics/43. Mike to provide words for potential new property for review.<br>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).<br>22/10: Mike wants to have a working implementation before closing on this, so marking the public comment as deferred.<br>31/10: Deferring for now |
| **241** | **Public comment: Bi-di properties placement in precedence section (All)**<br>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. |

## Work items:

| No | Item | Owner | Target | Status |
|---|---|---|---|---|
| 045 | Resolve public comments and incorporate into spec | All | 2013-10-22 | Pending |