

TeamRoom Plus

IBM TeamRoom

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

Meeting about Meetings\OGF

Project	DFDL 1.0
Meeting Date	23-Jul-13 (Tues)
Meeting Time	15:00 - 16:00

Created by Steve Hanson on 09-Mar-11

Last Modified by Steve Hanson on 22-Jul-13

OGF DFDL Working Group Call , 23 July 2013

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. Throwing errors from DFDL annotations

Two weeks ago the WG agreed an errata to add the XPath fn:error() function to the DFDL subset. Mike has since proposed an alternative whereby the dfdl:assert message property takes an expression. Pros and cons to be discussed.

3. Assert questions

Should an assert be allowed to throw a fatal error that stops the parse?
Should the spec relax the restriction that an assert and a discriminator are not allowed on the same object?
Should the spec relax the restriction that only one discriminator is allowed per object?
Clarify some of the wording in the assert section (see email from Jonathan Cranford).

4. Regular expression questions

Agree wording to state regex conformance to Unicode Technical Standard #18.
What should happen if a regular expression uses a Unicode block or category to match against a string in a non-Unicode encoding?

5. Using testKind 'pattern' in an assert or discriminator .

In the description for testPattern it says if the data is not scannable "the results are not predictable". This wording was added by action 190 where it was agreed that because testKind 'pattern' was peeking into the data stream and could be on any object, there would be no restrictions on its use. Wording needs improving to reflect implications of encodingErrorPolicy property .

6. Result of no match for lengthKind 'pattern'

Clarify that when there is no match to the lengthPattern regex, it is NOT a processing error, but instead results in length 0.

7. XPath fn:trace function

Mike would like this supported by DFDL. Mike to justify its addition.

8. 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

Apologies

Minutes

Meeting closed

Next regular call

Tues 13th August 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

Actions raised at this meeting

No	Action

Current Actions :

No	Action
066	<p>Investigate format for defining test cases (All)</p> <p>25/11: IBM to see if it is possible to publish its test case format .</p> <p>04/12: no update</p> <p>...</p> <p>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'</p> <p>24/03: No progress</p> <p>03/03: Discussions have been taking place on the subset of tests that will be provided .</p> <p>10/03: work is progressing</p> <p>17/03: work is progressing</p> <p>31/03: work is progressing</p> <p>14/04: And XML test case format has been defined and is being tested.</p> <p>21/04: Schema for TDML defined. Need to define how this and the test cases will be made public</p> <p>05/05: Work still progressing</p> <p>12/05: Work still progressing</p> <p>02/06: Work still progressing on technical and legal considerations</p> <p>...</p> <p>25/08: Will chase to allow Daffodil access to test cases . The WG should define how implementation confirm that they 'conform to DFDL v1'</p> <p>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.</p> <p>08/09: IBM still progressing</p> <p>15/09: IBM still progressing, expect tests to be available within a few weeks</p> <p>22/09: IBM still progressing, expect tests to be available within a few weeks</p> <p>29/09: Test cases are being prepared.</p> <p>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.</p> <p>13/10: Still progressing</p> <p>10/11: Legal issues cleared, IBM in process of collecting 100 example test cases, ideally ones that fit the 'extended conformance' of NCSA Daffodil</p> <p>17/11: Work is progressing on verifying the test cases. It should be possible to distribute to the WG in 2 weeks.</p> <p>24/11: About half the test cases have been completed and are being reviewed internally .</p> <p>01/12: Test cases should be available shortly</p> <p>08/12: The test cases are in internal IBM review. Probably need a bit of reorganising before publication</p> <p>Stephanie gave a brief overview of the format of the test cases.</p> <p>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.</p> <p>22/12: Test cases were sent to Joe for initial testing which found some problems in the Daffodil parser</p> <p>12/01: All current tests use a default format whih Daffodil doesn't currently support. Joe</p>

	<p>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.</p> <p>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.</p> <p>26/01: Action kicked off within IBM. There was a brief discussion about naming and organisation of test cases but no preferences were expressed</p> <p>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 .</p> <p>09/02: Steve had previously sent a list of areas to be tested. Please review.</p> <p>23/02: Please review Steve's list of areas to be tested</p> <p>02/03: Alan had reviewed Steve's list and we went through his comments. Agreed there is no need for separate tests for the info set or for dfdl: property lists, unions etc but comment will be added that these should be exercised during property testing.</p> <p>09/03: Alan updated the test document. Need more introduction and perhaps adopting the OGF template.</p> <p>30/03. Ownership of test document passed to Steve. This action is merged with 112 and will cover all aspects of compliance suite .</p> <p>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.</p> <p>...</p> <p>10/07/2012: Discussed schemes to create interchangeable tests . Ideally need a DFDL defined error code per failure, in conjunction with specific inserts .</p> <p>...</p> <p>26/3/2013: Resurrecting deferred action.</p> <p>We have got to the point where it makes sense to converge the IBM DFDL and Daffodil variations of .tdml file.</p> <p>Steve to seek permission from IBM to make the list of IBM DFDL error messages available to DFDL WG.</p> <p>...</p> <p>24/5: No further progress.</p> <p>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 info set comparison easier , Mark will see whether the shared tests use the latest version .</p> <p>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.</p> <p>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.</p> <p>...</p> <p>23/7: No further progress</p>
123	<p>DFDL tutorial (Steve)</p> <p>13/10: Draft of first 3 chapters has been written and will be distributed to WG</p> <p>10/11: Posted to grid forge here (http://forge.gridforum.org/sf/go/doc16106?nav=1), work continuing at IBM to define a standard example-based chapter framework and to author additional chapters. Contributors welcome!</p> <p>17/11: Steve, Stephanie and Alan had a meeting to discuss the best structure for the tutorial</p>

and decide which examples to use throughout. The meeting raised more questions. Further discussions will be held.

24/11: The list of topics to be covered in the remaining lessons has been produced and a lesson template. Alan will write lesson 4

01/12: Alan has started lesson 4 which covers fixed and variable fields and arrays.

08/12: Alan has almost completed lesson 4. Will send out for review.

15/12: First draft of lesson 4 is available for review. Alan to send to Bob and Joe.

22/12: Alan has distributed drafts for tutorials on Basic Structure and Optional/Repeating elements. Please review

12/01: Alan distributed a tutorial for choices and updated the others. Alan and Steve reviewed them and updated versions will be sent soon. Should start on the 'representation' tutorials soon.

19/01: The tutorials for basic structure, optional/arrays and choices have been updated. Please review. The tutorial for text elements should be available soon.

26/01: No comments received about 3 tutorials distributed last week. Alan is still working on Text representation.

02/02: Steve has sent comments on three tutorials. Alan to send updated versions by the end of the week. Alan has also distributed the first part of the tutorial on text representation and would like feedback.

09/02: Steve had reviewed tutorials 3,4,5 and updated versions have been distributed. Joe reviewed lesson on text elements.

Main points. Using 'represented as text' is confusing. Examples are too cluttered. Suggest simple targeted examples but still build up to final complete schema

23/02: New versions distributed and Steve has commented.

02/03: Alan has published the final versions of tutorials 4,5,6 and is working on text representations. There was some discussion about the detail that needs to be covered. Should limit it to 'common usage' and refer to the spec for details of edge cases.

09/03: Alan distributed an update to the text tutorial. Please review.

30/03: Steve has spent half a day tidying up lessons 1 to 6 and has uploaded them as pdfs to gridforge. They are now more coherent, and many inconsistencies and errors fixed. Ownership of draft lessons (text properties, binary properties, advanced features) has been passed to Steve. Also need to make a schema available for the examples.

13/04: Steve is working on the text properties tutorial.

04/05: No progress

...

15/06: This is on hold until Steve clears up spec issues and other workload. Steph has looked at the later lessons, and noted that they are more direct compared to the more wordy earlier lessons.

28/06: On hold.

...

29/11: Tim offered to take a look at the next outstanding tutorials. Steve / Tim to discuss

6/12: No progress

...

10/01: No progress, offer from Mike to help. First step is to make any corrections due to errata.

17/01: No progress

24/01: No update

31/01: Daffodil project team will be working their way through the existing tutorials and reviewing

14/02: Daffodil team to start reviewing tutorials hopefully this Friday.

21/02: Moved to this coming Friday

28/02: No update

13/03: No progress

21/03: No progress from Daffodil team. IBMers are starting to use the tutorial and will feedback any comments.

28/03: No change

05/04: Steve will send Alan's two draft lessons on binary & text data to Mike to complete.

17/04: No progress

	<p>8/5: No update</p> <p>...</p> <p>4/9: No progress:</p> <p>11/9: IBM DFDL infocenter will start to reference these directly before the end of the year , so they need updating soon.</p> <p>18/9: Noted that several requests have been received asking for chapters 7 to 17 as implied by chapter 1. At minimum chapter 1 needs updating to make it clear what exists today .</p> <p>28/9: Steve has updated and re-issued chapters 1 to 3.</p> <p>...</p> <p>12/2: No further progress</p> <p>19/2: Noted that tutorials need updating to reflect updated spec when it is issued .</p> <p>26/2: MITRE are using DFDL heavily now and suggesting ideas for tutorials .</p> <p>...</p> <p>9/7: No further progress</p> <p>16/7: Noted that at some point the lack of material will start to inhibit take up of DFDL .</p> <p>Steve has been asked to do some video sessions for IBM developerWorks .</p> <p>Possibility of MITRE obtaining an intern to create some more of the tutorials .</p> <p>23/7:</p>
172	<p>Clarify how a DFDL string literal is matched against the data stream (Tim)</p> <p>23/5: Non-trivial algorithm, worth stating it in the spec.</p> <p>..</p> <p>25/7: No progress.</p> <p>31/7: Tim has been making notes but nothing written up formally . Will include treatment of %WSP*;</p> <p>...</p> <p>16/7: No further progress</p> <p>23/7: Tim has started work on this.</p>
199	<p>Review the unordered sequence rewrite -into-a-repeating-choice section (Tim)</p> <p>16/1: This needs to be worded in a way that allows implementations to check for occurs violations in a manner which is consistent with the description of array processing in action 140 document, and is not overly constraining on implementers .</p> <p>22/1: Tim will propose revised words as he proceeds with the IBM implementation</p> <p>...</p> <p>26/2: No progress</p> <p>5/3: Tim has rewritten this and sent for review . Mike is happy with content, Steve to review.</p> <p>12/3: Reviewed an updated document from Tim that includes empty representation discussion . Tim to update further.</p> <p>...</p> <p>4/6: No further progress</p> <p>18/6: Tim has mailed the unordered rewrite for review . Steve & Mike have reviewed . Proposal to allow only dfdl:occursCountKind 'parsed' on optional and array elements in an unordered group. Also proposal that a flag is added to the infoset to indicate whether a value was defaulted, enabling implementers of the unordered rewrite algorithm to remove unwanted defaults at the end of the algorithm.</p> <p>9/7: Subsequent questions on whether occursCountKind is therefore irrelevant, and what is the behaviour for (1,1) elements. Agreed that we should try and keep the existing 'ordered' concepts where possible, so occursCountKind is needed for optional / array elements, and exactly one occurrence of a (1,1) element is expected, in the usual way. Tim to update the rewrite wording and send out for final review.</p> <p>16/7: Tim has updated the wording and re-issued. Comments from Mike and Steve. Tim to incorporate.</p> <p>Discussed where to place the extra constraints on XPath expressions that apply to unordered and choice . Related to the existing description in section 23.1. New action 214 created to resolve the XPath question.</p> <p>23/7: Tim has updated the wording, Steve has reviewed.</p>
200	<p>Establish recommended practices for pushing changes to GitHub (Mike)</p>

	<p>29/1: Mike will talk to Tresys who have used Git a lot.</p> <p>5/2: Mike to talk to Tresys this week, Tim has sent some links.</p> <p>12/2: Information sent by Mike, Steve to review.</p> <p>...</p> <p>23/7: No further progress</p>
204	<p>Establish strict versus lax behaviour for ICU calendar patterns (Steve)</p> <p>5/2: ICU ticket raised, response awaited</p> <p>12/2: Response received. The only leniency involved is when lax is specified, in which values can be outside the usual range and cause the calendar to be normalized, eg, 2013-01-35 would give 2013-02-04.</p> <p>19/2: Re-opened. More info from ICU. There is more leniency than published when lax, and there is also some leniency when strict. ICU will update the ticket with more details within 2 weeks.</p> <p>26/2: No update from ICU yet</p> <p>5/3: ICU working on this now</p> <p>12/3: Seems like it is a bigger job than ICU anticipated! They are still working on it.</p> <p>Steve has noted that IBM DFDL accepts any number of fractional seconds for 'SSS', not sure whether this is ICU leniency or mis-interpretation of the spec.</p> <p>19/3: No reply so far</p> <p>26/3: Reply received which describes the scope of leniency and what lenient behaviour is included in strict mode. Some of the information requires clarification so Steve has asked some further questions in the ticket.</p> <p>5/4: No further response yet.</p> <p>23/4: Further reply received from ICU. Steve will assemble into a coherent form and mail to the WG. IBM have proposed to ICU that they implement a 'super strict' mode where there is no leniency at all. DFDL could then expose this as a new enum 'exact' (say) when it appears in an ICU release. Same for text numbers.</p> <p>...</p> <p>18/6: Still with Steve</p> <p>9/7: Steve has a draft written but testing has uncovered some ICU4J bugs (EEEEEE and eee+) and some behaviour that does not match the ticket. Steve has asked for clarification via the ticket.</p> <p>16/7: No update</p> <p>23/7:</p>
205	<p>Establish XML white space rules for DFDL annotations (Suman)</p> <p>19/2: Steve will experiment with IBM DFDL and see what XML rules are being applied by default, and the effect of xml:space.</p> <p>26/2: Using IBM DFDL, dfdl:assert has external white space (ie, outside the { }) stripped, and internal white space (ie, inside the { }) normalised to a single space, for both attribute and element renderings. Use of xml:space 'preserve' made no difference. Also looked at DFDL properties such as dfdl:length '10' and dfdl:initiator 'abc def' and they showed the same behaviour. However enum properties like dfdl:lengthKind 'explicit' give a schema definition error. Is this deliberate? Action passed to Suman to confirm the intended behaviour.</p> <p>5/3: IBM DFDL uses JAXP which handles white space according to type. In Suman's XSDs for DFDL enums are modelled as xs:string (spaces preserved), but DFDL string literals are modelled as xs:token (spaces trimmed and collapsed). This explains the observed behavior. Steve will update section 6.3 to make clear the behaviour expected for the different property types and send for review.</p> <p>12/3: No progress</p> <p>19/3: Further investigation: DFDL integer properties get leading/trailing spaces stripped. DFDL string will collapse all white space, not just leading/trailing. For DFDL expressions DFDL should use xs:string, but trim off leading and trailing white space, which it is safe to do. Need to see how pattern facets are handled and base DFDL regular expression on that.</p> <p>26/3: The XSD schema-for-schemas has pattern facet as xs:string, which implies DFDL should do same for regexs. Also noted that the whitespace facet (an enum) is xs:NMTOKEN which derives from xs:token. Does that change what we had decided about DFDL enums? Steve to</p>

	<p>investigate further.</p> <p>...</p> <p>18/6: Still with Steve</p> <p>9/7: Steve's recommendation is;</p> <ul style="list-style-type: none"> - Enumeration changes from xs:string to xs:token (reason: to match XSDL more closely and trim leading/trailing whitespace) - DFDL regular expression stays as xs:string (reason: regex may contain literal white space) - DFDL string literal changes from xs:string to xs:token (reason: currently inconsistent with List of DFDL string literal) - List of DFDL string literal stays as list of xs:token - DFDL expression changes from xs:token to xs:string (reason: XPath may contain non-ignorable whitespace) <p>Further:</p> <ul style="list-style-type: none"> - DFDL regular expression should not trim leading/trailing whitespace - DFDL expression should trim leading whitespace before { and trailing whitespace after } - The enum of DFDL property names should be based on xs:token <p>Noted that xs:NMTOKEN is intended for attributes only, so should not be used for DFDL properties as they can be expressed in attribute or element forms.</p> <p>Need to get Suman's approval for the proposed changes.</p> <p>16/7: Discussed the nature of DFDL string literals, noted that DFDL entities are not the same as XML entities.</p> <p>Mike pointed out that xs:token on its own is not sufficient for DFDL string literal as it collapses internal whitespace, rather than disallowing it. Steve to investigate whether xs:NMTOKEN is a better match or whether a pattern facet is needed.</p> <p>23/7: Steve thinks xs:token plus a pattern that disallows a space is the best model.</p>
208	<p>Create errata v 13 and incorporate into DFDL spec (Steve, Mike)</p> <p>26/3: Mike to review Steve's errata v13 document.</p> <p>5/4: Looks ok, Steve will publish on redmine.</p> <p>23/4: Published v13 of the errata document.</p> <p>7/5: One TBD in v13 for Tim's 'unordered rewrite' - outlook is a couple of weeks. Steve will add in recent errata if he has time.</p> <p>24/5: Tim's unordered rewrite not ready, Steve not added most recent errata</p> <p>28/5: No progress</p> <p>4/6: No progress</p> <p>18/6: When the unordered rewrite document is agreed, Steve will incorporate into the errata document.</p> <p>9/7: MITRE are asking when the next official draft of the spec will be available. Confirm date next week.</p> <p>16/7: Mike has incorporated errata v 13 and sent out spec draft r13. Steve to review.</p> <p>23/7: Steve has reviewed and sent comments.</p>
213	<p>New DFDL functions to create numbers from hex strings (Mike)</p> <p>18/6: No way in XPath to cast a literal hex constant to a Number type, need DFDL functions if we want to do this, Mike to send proposal.</p> <p>9/7: Mike has sent proposal for a set of new Number constructors that match the existing Number constructors except they also allow an 'x' followed by a hex string as argument, eg, dfdl:short('x12AB'). Agreed on the proposal. Mike also proposed an inverse dfdl:hex() function. Questions about the name, whether 'x' should appear, and endian-ness. Mike to revise proposal.</p> <p>16/7: Revised proposal sent for dfdl:hexBinary function. Got somewhat lost in the fn:error discussion, Mike to resend.</p> <p>23/7: Proposal resent, Tim and Steve have reviewed.</p>
214	<p>Expression Language Data Model (All)</p> <p>16/7: Augment section 23.1 so that it covers the cases where an XPath references an element that is in a different choice branch or that is in the same unordered sequence or that is floating. These cases could be detected statically (though to do this 100% reliably is not easy) or they could be left until runtime and fail if the element does not exist. Both are schema definition</p>

errors as explained by errata 2.120.
[23/7:](#)

Closed actions

No	Action

Deferred actions

No	Action
129	Press release to publicise DFDL (Steve) Steve is pulling together a press release at IBM. Want to include as many contributors and interested parties as possible. NCSA are keen to be included. Also likely that US National Archive will want to be included. Mike has indicated OCO are too. 17/11: no progress ... 08/12: Still no response from IBM press office 15/12: no progress 09/03: No progress 30/03: Making this action deferred until IBM is in a position to say something more concrete about any implementation.
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 gridforge 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

Work items:

No	Item	Owner	Target	Status
043	Track errata list for 1.0 of the spec.	Steve	N/A	Draft 013 on Redmine.
044	Incorporate errata list into DFDL spec.	Steve/Mike	N/A	Draft merged document.