

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	13-Aug-13 (Tues)
Meeting Time	16:00 - 17:00

Created by Steve Hanson on 09-Mar-11

Last Modified by Steve Hanson on 14-Aug-13

OGF DFDL Working Group Call , 13 August 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. Clarify the relationship between validation errors and points of uncertainty

Tresys think the words in the spec on this matter are not as clear as they could be .

3. Possible mistake in erratum 3.1

This errata allows for dfdl:hiddenGroupRef to appear on choices.

4. Clarifications needed to errata 2.121

This revised erratum is not clear enough on what time zone symbols and variations are permitted by DFDL .

5. Disallow %WSP*; appearing on its own where %WSP*; is allowed but %ES; is not

On the grounds that %WSP*; on its own allows empty string.

6. Clarify the definitions of 'scan' and 'scannable'

The term 'scan' is used for the processing of the data stream when parsing for both lengthKind 'delimited' and 'pattern', but only the latter needs to be 'scannable'. This seems odd if you think of 'scannable' as meaning 'can be scanned'.

7. 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

Jonathan Cranford
Suman Kalia
Tim Kimber
Mike Beckerle

Apologies

Minutes

1. Daffodil Open Source Project

Nothing new to report.

2. Clarify the relationship between validation errors and points of uncertainty

Tresys to provide some improved words. New [action 216](#).

3. Possible mistake in erratum 3.1

This errata allows for `dfdl:hiddenGroupRef` to appear on choices. This was not deliberate, and was probably due to mixing up that the target hidden group could be a choice. The `hiddenGroupRef` property can only appear on an (empty) `xs:sequence`. [Errata 3.1 will be updated](#).

4. Clarifications needed to errata 2.121

Agreed with Steve's clarifications in his email, namely that the new symbols 'x', 'X' and 'O' will be supported (all variations) and the existing symbol 'V' will adopt the new semantics (all variations). [Errata 2.121 will be updated](#).

5. Disallow `%WSP*`; appearing on its own where `%WSP*`; is allowed but `%ES`; is not

Agreed that it should be. [Errata taken](#).

6. Clarify the definitions of 'scan' and 'scannable'

Agreed that the terms 'scan' and 'scannable', as currently defined, are not clear. The current definition of 'scan' will be retained and all derived words such as 'scanning' and 'scannable' are to be consistent with this. A new term will be introduced to describe the more stringent requirements for regular expression scanning, suggestions included 'scannable-as-text' or 'matchable'.

[New action 217](#).

Meeting closed

17:15 UK

Next regular call

Tues 20th August 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

Subject	Document Type	Created	Modified

Next action: 218

Actions raised at this meeting

No	Action
216	Clarify the relationship between validation errors and points of uncertainty (Mike) 13/8: Tresys to provide some improved words.
217	New term for to replace 'scannable' (Mike) 13/8: Decide on a new phrase for the regex scanning requirement, such as 'scannable-as-text' or 'matchable'.

Current Actions :

No	Action
066	Investigate format for defining test cases (All) 25/11:IBM to see if it is possible to publish its test case format . 04/12: no update ... 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' 24/03: No progress 03/03: Discussions have been taking place on the subset of tests that will be provided . 10/03: work is progressing 17/03: work is progressing 31/03: work is progressing 14/04: And XML test case format has been defined and is being tested. 21/04. Schema for TDML defined. Need to define how this and the test cases will be made

public

05/05: Work still progressing

12/05: Work still progressing

02/06: Work still progressing on technical and legal considerations

...

25/08: Will chase to allow Daffodil access to test cases. The WG should define how implementation confirm that they 'conform to DFDL v1'

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.

08/09: IBM still progressing

15/09: IBM still progressing, expect tests to be available within a few weeks

22/09: IBM still progressing, expect tests to be available within a few weeks

29/09: Test cases are being prepared.

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.

13/10: Still progressing

10/11: Legal issues cleared, IBM in process of collecting 100 example test cases, ideally ones that fit the 'extended conformance' of NCSA Daffodil

17/11: Work is progressing on verifying the test cases. It should be possible to distribute to the WG in 2 weeks.

24/11: About half the test cases have been completed and are being reviewed internally.

01/12: Test cases should be available shortly

08/12: The test cases are in internal IBM review. Probably need a bit of reorganising before publication

Stephanie gave a brief overview of the format of the test cases.

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.

22/12: Test cases were sent to Joe for initial testing which found some problems in the Daffodil parser

12/01: All current tests use a default format which 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.

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.

26/01: Action kicked off within IBM. There was a brief discussion about 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.

...

13/8: No further progress

123

DFDL tutorial (Steve)

13/10: Draft of first 3 chapters has been written and will be distributed to WG

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!

17/11: Steve, Stephanie and Alan had a meeting to discuss the best structure for the tutorial 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
8/5: No update
...
4/9: No progress:
11/9: IBM DFDL infocenter will start to reference these directly before the end of the year, so they need updating soon.
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.
28/9: Steve has updated and re-issued chapters 1 to 3.
...
12/2: No further progress
19/2: Noted that tutorials need updating to reflect updated spec when it is issued.
26/2: MITRE are using DFDL heavily now and suggesting ideas for tutorials.
...
9/7: No further progress
16/7: Noted that at some point the lack of material will start to inhibit take up of DFDL. Steve has been asked to do some video sessions for IBM developerWorks. Possibility of MITRE obtaining an intern to create some more of the tutorials.
...
23/7: No updates
13/8: Will need to be updated to match forthcoming spec draft. At same time decide on what lessons are next in priority.

172 **Clarify how a DFDL string literal is matched against the data stream (Tim)**

23/5: Non-trivial algorithm, worth stating it in the spec.

..
25/7: No progress.

	<p>31/7: Tim has been making notes but nothing written up formally. Will include treatment of %WSP*; ... 16/7: No further progress 23/7: Tim has started work on this. Scope is the matching of a single DFDL String Literal. 13/8: No further progress</p>
200	<p>Establish recommended practices for pushing changes to GitHub (Mike) 29/1: Mike will talk to Tresys who have used Git a lot. 5/2: Mike to talk to Tresys this week, Tim has sent some links. 12/2: Information sent by Mike, Steve to review. ... 13/8: No further progress</p>
204	<p>Establish strict versus lax behaviour for ICU calendar patterns (Steve) 5/2: ICU ticket raised, response awaited 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. 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. 26/2: No update from ICU yet 5/3: ICU working on this now 12/3: Seems like it is a bigger job than ICU anticipated! They are still working on it. 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. 19/3: No reply so far 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. 5/4: No further response yet. 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. ... 18/6: Still with Steve 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. 16/7: No update 23/7: Clarification ongoing 13/8: Steve to write up.</p>
215	<p>Create errata v 14 and incorporate into DFDL spec (Steve, Mike) 23/7: Steve has created a draft including all errata up until 9th July, rest to be added. 13/8: Steve has completed errata v014. Mike has a draft r14 of the spec that includes the errata. Steve has reviewed and commented, some of the comments required discussion, and this is documented in the email thread. A separate call will be held to complete the review, and after that Mike will update draft r14 of the spec. Noted that there are still other OPEN comments in the spec which need resolving.</p>

Closed actions

No	Action
208	<p>Create errata v 13 and incorporate into DFDL spec (Steve, Mike) 26/3: Mike to review Steve's errata v13 document. 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 v13 and sent out spec draft r13. Steve to review.</p> <p>23/7: Steve has reviewed and sent comments. Mike to update spec draft r13, includes minor editorial comments that have built up. Any further significant work to be done in draft r14. See new action 215.</p> <p>13/8: Closed. Mike has completed draft r13.</p>

Deferred actions

No	Action
129	<p>Press release to publicise DFDL (Steve)</p> <p>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.</p> <p>17/11: no progress</p> <p>...</p> <p>08/12: Still no response from IBM press office</p> <p>15/12: no progress</p> <p>....</p> <p>09/03: No progress</p> <p>30/03: Making this action deferred until IBM is in a position to say something more concrete about any implementation.</p>
131	<p>Transformation of DFDL properties to a canonical form (Joe)</p> <p>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.</p> <p>15/12: Alan tested against test dfdl schema which worked correctly (after fixing some errors in the schema)</p> <p>22/12: no update</p> <p>12/01: Joe has some defects to fix before making available on gridforge.</p> <p>19/01: There is a difficult problem to solve before Joe make the style sheet public</p> <p>26/01: Working on problems</p> <p>02/02: no progress</p> <p>09/02: As it wasn't a simple as exoected this will be treated as a low priority action</p> <p>23/02: Low priority</p> <p>09/03: Low priority</p> <p>30/03: Deferring for now</p>

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.

