

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	03-Sep-13 (Tues)
Meeting Time	17:00 - 18:00

Created by Steve Hanson on 09-Mar-11

Last Modified by Steve Hanson on 05-Sep-13

OGF DFDL Working Group Call , 3 September 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. Standardizing on text /textual

Steve favours dropping 'textual' and using the phrases 'text representation' and 'binary representation' consistently throughout.

3. Documenting fallback behaviour for timezones

For certain time zone symbols, ICU prescribes a fallback if there is no data for that symbol in a particular zone . DFDL spec needs to document this somehow. Can it be disabled?

4. 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
Mike Beckerle
Suman Kalia
Jonathan Cranford

Apologies

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. Standardizing on text /textual

Use of 'textual' ok in early discussion of formats. Later in the spec we should use glossary-defined terms 'text representation' and 'binary representation' to avoid the clumsy 'dfdl:representation "text" or type xs:string' and 'dfdl:representation "binary" or type xs:hexBinary' phrases, respectively. Could still use 'textual' to describe a text oriented format, ie, text data and delimiters? **New action 225** raised.

3. Documenting fallback behaviour for timezones

Steve will find obtain the description of the fallback behavior from ICU . **New action 223** raised.

4. Implementation defined limits .

All implementation defined limits should be grouped together for convenience . This has been started in section 2.6 but that is not really the correct place. Also Jonathan noted distinction between this and 'implementation dependent'. **New action 224** raised.

Meeting closed

18:00 UK

Next regular call

Tues 10th Sept 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: **226**

Actions raised at this meeting

No	Action
223	How best to document ICU time zone symbol fallbacks (Steve) 3/9: Steve to contact ICU. Resolve during public comment.
224	Add section for implementation defined limits (All) 3/9: Several places in the spec cite this, should be grouped. Currently partially listed in section 2.6. Also note distinction between 'implementation defined' and 'implementation dependent'. Check spec for correct usage. Resolve during public comment.
225	Consistent terminology for describing text data (All) 3/9: Ensure spec uses 'textual' and other agreed terms in a consistent way. Resolve during public comment.

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 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.
 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 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.

	<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 infoset 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>3/9: 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 and decide which examples to use throughout. The meeting raised more questions. Further discussions will be held.</p> <p>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</p> <p>01/12: Alan has started lesson 4 which covers fixed and variable fields and arrays .</p> <p>08/12: Alan has almost completed lesson 4. Will send out for review.</p> <p>15/12: First draft of lesson 4 is available for review. Alan to send to Bob and Joe.</p> <p>22/12: Alan has distributed drafts for tutorials on Basic Structure and Optional/Repeating elements. Please review</p> <p>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.</p> <p>19/01: The tutorials for basic structure, optional/arrays and choices have be updated. Please review. The tutorial for text elements should be available soon .</p> <p>26/01: No comments received about 3 tuorials distributed last week . Alan is still working on Text representation.</p> <p>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.</p>

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.

...

	<p>23/7: No updates</p> <p>13/8: Will need to be updated to match forthcoming spec draft. At same time decide on what lessons are next in priority.</p> <p>...</p> <p>28/8: No further progress</p> <p>3/9: Steve has updated lessons 1 through 5 to match the latest spec draft. Direct dispatch choice added to lesson 5. Lesson 1 lists the future lessons by number, we may want to revisit the proposed set at some point.</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. Scope is the matching of a single DFDL String Literal.</p> <p>13/8: No further progress</p> <p>20/8: Tim has circulated an initial draft. Review for next week, comments in advance via email please.</p> <p>28/8: Tim will incorporate Mike's comments and re-issue.</p> <p>3/9: Added to spec draft r14.4 as an appendix. Steve has reviewed, comments sent. Comment 2 to be addressed now, comments 1 and 3 to be handled in public comment.</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>3/9: No further progress</p>
215	<p>Create errata v 14 and incorporate into DFDL spec (Steve, Mike)</p> <p>23/7: Steve has created a draft including all errata up until 9th July, rest to be added.</p> <p>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> <p>20/8: Separate call held to complete review. A couple of issues outstanding which will be handled as separate agenda items.</p> <p>Mike has mailed out draft r14.2 of the spec and updated the errata document to v014.2. Mike has listed the remaining items to complete. Keep this action open until all work complete and ready for distribution.</p> <p>28/8: Mike has mailed out errata v014.3, spec draft r14.3 and a list of open items. Steve has reviewed and sent comments.</p> <p>3/9: Latest is errata v014.5 and spec draft r14.5. Mike to create draft r14.6 containing last few content changes and send to WG for proof reading. There will be a draft r14.7 for proof reading comments and other editorial changes. Prior to public comment, all format changes will be accepted and all comments deleted. Change bars will remain. A fully commented copy will be retained on Redmine.</p>
220	<p>Review reworked information on bit fields (Steve)</p> <p>20/8: Major changes to presentation rather than content, but a full review is needed even so.</p> <p>28/8: Mike and Steve collaborating on getting this into the spec and reviewed. Spotted an issue in passing with the grammar, the RightPadOrFill region could actually contain pad characters and fill bytes, for the text & lengthUnits 'bytes' combination. Erratum taken.</p> <p>3/9: Draft r14.5 contains updates. Went through the remaining open comments. Most can be closed now, a few remain and will be handled in public comment, specifically</p> <ul style="list-style-type: none"> - adding lengthUnits 'bits' for packed calendars;

	- documenting the endOfParent case where RightFill can be encountered.
221	Ensure 'nil' is handled by DFDL expressions (All) 20/8: Possible problem with 'nil' and expressions. Needs investigation. 28/8: No further progress. Unlikely to make the published draft. 3/9: No progress - deferred until public comment
222	Improve some of the wording used in separator suppression description (All) 28/8: Mike has some concerns with the accuracy of the wording, see email thread. 3/9: No progress - deferred until public comment

Closed actions

No	Action
216	Clarify the relationship between validation errors and points of uncertainty (Mike) 13/8: Tresys to provide some improved words. 20/8: No progress 28/8: Words added into spec draft r14.3. Pending review by Steve. 3/9: Closed. Comment sent via email.
217	New term to replace 'scannable' (Mike) 13/8: Decide on a new phrase for the regex scanning requirement, such as 'scannable-as-text' or 'matchable'. 20/8: No progress 28/8: Mike has reservations (see email thread) and thinks that reverting the behaviour back to what was originally prescribed in the GA spec is a better way forward, given that we now have the encodingErrorPolicy property which has cleared up the behaviour when decoding goes wrong. So lengthKind 'pattern' would behave like testKind 'pattern' in an assert or discriminator, specifically, encoding must be in scope and there are no further rules that apply to the element nor to children of a complex element. It is thus allowed on binary elements. The equivalent rules for a complex element with specified length and lengthUnits 'characters' should be dropped, for consistency. Erratum 3.9 updated. 3/9: Closed. Draft r14.5 contains these updates. Proof readers please check for consistency.
219	Decide on revised rules for direct dispatch choice (All) 20/8: Need to satisfy all known use cases with good performance, but at the same time allow QNames in a future DFDL release. 28/8: Proposal sent via email, the dispatch mechanism becomes entirely local to a choice. Need to agree on revised property names. 3/9: Closed. Agreed on choiceDispatchKey and choiceBranchKey.

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 gidforge 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.