ISO/IEC 23415:2024-04 (E)

Information technology - Data Format Description Language (DFDL) v1.0 Specification

Contents

Page

1 Introduction 9
1.1 Why is DFDL Needed? 10
1.2 What is DFDL? 10
1.2.1 Simple Example 10
1.3 What DFDL is not 13
1.4 Scope of version 1.0 13
2 Overview of the Specification 15
3 Notational and Definitional Conventions 16
3.1 Glossary and Terminology 16
3.2 Failure Types 16
4 The DFDL Information Set (Infoset) 17
4.1 "No Value" 18
4.2 Information Items 18
4.2.1 Document Information Item 18
4.2.2 Element Information Items 18
4.3 DFDL Information Item Order 19
4.4 DFDL Augmented Infoset 19
5 DFDL Schema Component Model 20
5.1 DFDL Simple Types 20
5.2 DFDL Subset of XML Schema 21
5.3 XSD Facets, min/maxOccurs, default, and fixed 22
5.3.1 MinOccurs, MaxOccurs 23
5.3.2 MinLength, MaxLength 23
5.3.3 MaxInclusive, MaxExclusive, MinExclusive, MinInclusive, TotalDigits, FractionDigits 23
5.3.4 Pattern 23
5.3.5 Enumeration Values 23
5.3.6 Default 23
5.3.7 Fixed 24
5.4 Compatibility with Other Annotation Language Schemas 24
6 DFDL Syntax Basics 25
6.1 Namespaces 25
6.2 The DFDL Annotation Elements 25
6.3 DFDL Properties 26
6.3.1 DFDL String Literals 28
6.3.2 DFDL Expressions 32
6.3.3 DFDL Regular Expressions 32
6.3.4 Enumerations in DFDL 32
7 Syntax of DFDL Annotation Elements 33
7.1 Component Format Annotations 33
7.1.1 Property Binding Syntax 34
7.1.2 Empty String as a Representation Property Value 35
7.2 dfdl:defineFormat - Reusable Data Format Definitions 36

7.2.1 Using/Referencing a Named Format Definition: The dfdl:ref Property 36
7.2.2 Inheritance for dfdl:defineFormat 36
7.3 The dfdl:defineEscapeScheme Defining Annotation Element 37
7.3.1 Using/Referencing a Named escapeScheme Definition 37
7.4 The dfdl:escapeScheme Annotation Element 37
7.5 The dfdl:assert Statement Annotation Element 38
7.5.1 Properties for dfdl:assert 38
7.6 The dfdl:discriminator Statement Annotation Element 40
7.6.1 Properties for dfdl:discriminator 40
7.7 DFDL Variable Annotations 43
7.7.1 dfdl:defineVariable Annotation Element 43
7.7.2 The dfdl:newVariableInstance Statement Annotation Element 44
7.7.3 The dfdl:setVariable Statement Annotation Element 45
8 Property Scoping and DFDL Schema Checking 47
8.1 Property Scoping 47
8.1.1 Property Scoping Rules 47
8.1.2 Providing Defaults for DFDL properties 47
8.1.3 Combining DFDL Representation Properties from a dfdl:defineFormat 48
8.1.4 Combining DFDL Properties from References 48
8.2 DFDL Schema Checking 51
8.2.1 Schema Component Constraint: Unique Particle Attribution 51
8.2.2 Optional Checks and Warnings 51
9 DFDL Processing Introduction 53
9.1 Parser Overview 53
9.1.1 Points of Uncertainty 53
9.1.2 Processing Error 54
9.1.3 Recoverable Error 54
9.2 DFDL Data Syntax Grammar 54
9.2.1 Nil Representation 56
9.2.2 Empty Representation 56
9.2.3 Normal Representation 57
9.2.4 Absent Representation 57
9.2.5 Zero-length Representation 57
9.2.6 Missing 57
9.2.7 Examples of Missing and Empty Representation 58
9.2.8 Round Trip Ambiguities 58
9.3 Parsing Algorithm 58
9.3.1 Known-to-exist and Known-not-to-exist 59
9.3.2 Establishing Representation 60
9.3.3 Resolving Points of Uncertainty 61
9.4 Element Defaults 62
9.4.1 Definitions 62
9.4.2 Element Defaults When Parsing 62
9.4.3 Element Defaults When Unparsing 64
9.5 Evaluation Order for Statement Annotations 65
9.5.1 Asserts and Discriminators with testKind 'expression' 66
9.5.2 Discriminators with testKind 'expression' 66
9.5.3 Elements and setVariable 66
9.5.4 Controlling the Order of Statement Evaluation 66
9.6 Validation 66
9.7 Unparser Infoset Augmentation Algorithm 67
10 Overview: Representation Properties and their Format Semantics 68

11 Properties Common to both Content and Framing 69
11.1 Unicode Byte Order Mark (BOM) 71
11.2 Character Encoding and Decoding Errors 71
11.2.1 Property dfdl:encodingErrorPolicy 72
11.2.2 Unicode UTF-16 Decoding/Encoding Non-Errors 73
11.2.3 Preserving Data Containing Decoding Errors 73
11.3 Byte Order and Bit Order 73
11.4 dfdl:bitOrder Example 73
11.4.1 Example Using Right-to-Left Display for 'leastSignificantBitFirst' 74
11.4.2 dfdl:bitOrder and Grammar Regions 74
12 Framing 75
12.1 Aligned Data 75
12.1.1 Implicit Alignment 76
12.1.2 Mandatory Alignment for Textual Data 76
12.1.3 Mandatory Alignment for Packed Decimal Data 77
12.1.4 Example: AlignmentFill 77
12.2 Properties for Specifying Delimiters 78
12.3 Properties for Specifying Lengths 81
12.3.1 dfdl:lengthKind 'explicit' 82
12.3.2 dfdl:lengthKind 'delimited' 82
12.3.3 dfdl:lengthKind 'implicit' 83
12.3.4 dfdl:lengthKind 'prefixed' 84
12.3.5 dfdl:lengthKind 'pattern' 87
12.3.6 dfdl:lengthKind 'endOfParent' 87
12.3.7 Elements of Specified Length 89
13 Simple Types 92
13.1 Properties Common to All Simple Types 92
13.2 Properties Common to All Simple Types with Text representation 93
13.2.1 The dfdl:escapeScheme Properties 94
13.3 Properties for Bidirectional support for All Simple Types with Text representation 97
13.4 Properties Specific to String 97
13.5 Properties Specific to Number with Text or Binary Representation 99
13.6 Properties Specific to Number with Text Representation 99
13.6.1 The dfdl:textNumberPattern Property 106
13.6.2 Converting logical numbers to/from text representation 110
13.7 Properties Specific to Number with Binary Representation 112
13.7.1 Converting Logical Numbers to/from Binary Representation 113
13.8 Properties Specific to Float/Double with Binary Representation 118
13.9 Properties Specific to Boolean with Text Representation 118
13.10 Properties Specific to Boolean with Binary Representation 119
13.11 Properties Specific to Calendar with Text or Binary Representation 120
13.11.1 The dfdl:calendarPattern property 122
13.11.2 The dfdl:calendarCheckPolicy Property 125
13.12 Properties Specific to Calendar with Text Representation 125
13.13 Properties Specific to Calendar with Binary Representation 126
13.14 Properties Specific to Opaque Types (xs:hexBinary) 127
13.15 Nil Value Processing 127
13.16 Properties for Nillable Elements 127

14 Sequence Groups 131
14.1 Empty Sequences 131
14.2 Sequence Groups with Separators 132
14.2.1 Separators and Suppression 133
14.2.2 Parsing Sequence Groups with Separators 134
14.2.3 Unparsing Sequence Groups with Separators 136
14.3 Unordered Sequence Groups 138
14.3.1 Restrictions for Unordered Sequences 138
14.3.2 Parsing an Unordered Sequence 138
14.3.3 Unparsing an Unordered Sequence 140
14.4 Floating Elements 140
14.5 Hidden Groups 141
15 Choice Groups 143
15.1 Resolving Choices 144
15.1.1 Resolving Choices via Speculation 144
15.1.2 Resolving Choices via Direct Dispatch 145
15.1.3 Unparsing Choices 145
16 Properties for Array Elements and Optional Elements 146
16.1 The dfdl:occursCountKind property 146
16.1.1 dfdl:occursCountKind 'fixed' 146
16.1.2 dfdl:occursCountKind 'implicit' 147
16.1.3 dfdl:occursCountKind 'parsed' 147
16.1.4 dfdl:occursCountKind 'expression' 147
16.1.5 dfdl:occursCountKind 'stopValue' 147
16.2 Default Values for Arrays 148
16.3 Arrays with DFDL Expressions 148
16.4 Points of Uncertainty 148
16.5 Arrays and Sequences 148
16.6 Forward Progress Requirement 148
16.7 Parsing Occurrences with Non-Normal Representation 149
16.8 Sparse Arrays 149
17 Calculated Value Properties 150
17.1 Example: 2d Nested Array 151
17.2 Example: Three-Byte Date 151
18 DFDL Expression Language 154
18.1 Expression Language Data Model 154
18.2 Variables 155
18.2.1 Rewinding of Variable Memory State 155
18.2.2 Variable Memory State Transitions 155
18.3 General Syntax 156
18.4 DFDL Expression Syntax 157
18.5 Constructors, Functions and Operators 158
18.5.1 Constructor Functions for XML Schema Built-in Types 158
18.5.2 Standard XPath Functions 159
18.5.3 DFDL Functions 162
18.5.4 DFDL Constructor Functions 164
18.5.5 Miscellaneous Functions 165
18.6 Unparsing and Circular Expression Deadlock Errors 166

19 DFDL Regular Expressions 167
20 External Control of the DFDL Processor 168
21 Built-in Specifications 169
22 Conformance 170
23 Optional DFDL Features 171
24 Security Considerations 173
25 Authors and Contributors 174
26 Intellectual Property Statement 175
27 Disclaimer 176
28 Full Copyright Notice 177
29 References 178
30 Appendix A: Escape Scheme Use Cases 181
30.1 Escape Character Same as dfdl:escapeEscapeCharacter 181
30.2 Escape Character Different from dfdl:escapeEscapeCharacter 181
30.2.1 Example 1 - Separator ';' 181
30.2.2 Example 2 - Separator 'sep' 182
30.3 Escape Block with Different Start and End Characters 182
30.4 Escape Block with Same Start and End Characters 183
31 Appendix B: Rationale for Single-Assignment Variables 185
32 Appendix C: Processing of DFDL String literals 186
32.1 Interpreting a DFDL String Literal 186
32.2 Recognizing a DFDL String Literal 186
32.3 Recognizing DFDL String Literal Part 186
33 Appendix D: DFDL Standard Encodings 188
33.1 Purpose 188
33.2 Conventions 188
33.3 Specification Template 188
33.4 Encoding X-DFDL-US-ASCII-7-BIT-PACKED 188
33.4.1 Name 188
33.4.2 Translation table 188
33.4.3 Width 188
33.4.4 Alignment 189
33.4.5 Byte Order 189
33.4.6 Example 1 189
33.4.7 Example 2 189
33.5 Encoding X-DFDL-US-ASCII-6-BIT-PACKED 191
33.5.1 Name 191
33.5.2 Translation Table 191
33.5.3 Width 192
33.5.4 Alignment 192
33.5.5 ByteOrder 192
33.5.6 Example 1 192
33.6 References for Appendix D 193
34 Appendix E: Glossary of Terms 194
35 Appendix F: Specific Errors Classified 200
36 Appendix G: Property Precedence 202
36.1 Parsing 202
36.1.1 dfdl:element (simple) and dfdl:simpleType 202

36.1.2 dfdl:element (complex) 205
36.1.3 dfdl:sequence and dfdl:group (when reference is to a sequence) 206
36.1.4 dfdl:choice and dfdl:group (when reference is to a choice) 207
36.2 Unparsing 208
36.2.1 dfdl:element (simple) and dfdl:simpleType 208
36.2.2 dfdl:element (complex) 212
36.2.3 dfdl:sequence and dfdl:group (when reference is a sequence) 213
36.2.4 dfdl:choice and dfdl:group (when reference is a choice) 213