PAS Explanatory Report for Open Grid Forum (OGF) Data Format Description Language (DFDL)

ISO/IEC JTC 1 Common Strategic Characteristics

PAS Originators/Submitters are invited to explicitly reference the JTC 1 common strategic characteristics (interoperability, portability, cultural and linguistic adaptability, and accessibility) when submitting their PAS Submitter application or any PAS for transposition.

ORGANIZATION CRITERIA (SD9 7.3)

OGF was approved as a PAS Submitter effective 2022-12-02; the application ID is N16164.
Since attaining PAS Submitter status there have been no significant changes to OGF at an organizational level. Additionally, there have been no changes to the organization criteria as submitted in the PAS submitter application N16164.

DOCUMENT RELATED CRITERIA (SD9 7.4)

The DFDL 1.0 Specification provides a modeling language for describing general text and binary data in a standard way. A DFDL model or schema allows any text or binary data to be read (or "parsed") from its native format and to be presented as an instance of an information set. (An information set is a logical representation of the data contents, independent of the physical format. For example, two records could be in different formats, because one has fixed-length fields and the other uses delimiters, but they could contain exactly the same data, and would both be represented by the same information set). The same DFDL schema also allows data to be taken from an instance of an information set and written out (or "serialized") to its native format.

This specification addresses a specific challenge faced in response to a need for interoperable APIs to be able to understand data regardless of source. A language was needed capable of modeling a wide variety of existing text and binary data formats. A working group was established at the Global Grid Forum (which later became the Open Grid Forum) in 2003 to create a specification for such a language which has now become a Full Recommendation of the OGF.

This submission addresses the document related criteria specified in SD 9 clause 7.4 as follows:

7.4.1 Quality

Within its scope the specification shall completely describe the functionality (in terms of interfaces, protocols, formats, etc.) necessary for an implementation of the submission. If it is based on a product, it shall include all the functionality necessary to achieve the stated level of compatibility or interoperability in a product independent manner.

7.4.1.1 Completeness (M)

a) How well are all interfaces specified?

The DFDL 1.0 Specification has multiple implementations including several different implementations by IBM Corporation, DFDL4S by the European Space Agency, and an open-source implementation (Apache Daffodil). Interoperability of the IBM Java-based DFDL implementation and the Apache Daffodil implementation has been demonstrated for several DFDL format descriptions (called 'schemas').

b) How easily can implementation take place without need of additional descriptions?

The specification is self-contained. It references other documents which collectively provide everything required to implement the language.

c) What proof exists for successful implementations (e.g. availability of test results for media standards)?

IBM has three independent DFDL implementations, one in Java, two in C, in use across these products:

- IBM WebSphere Message Broker (2011 - 2017)
- IBM Integration Bus (2013 - 2022)
- IBM App Connect Enterprise (2018 onwards)
- IBM z/Transaction Processing Facility (2014 onwards)
- IBM InfoSphere Master Data Management (2013 onwards)
- IBM Rational Test Workbench (2012 onwards)
- IBM Rational Integration Tester (2016 onwards)

The Apache Daffodil(tm) open-source implementation has been accepted as an Apache Software Foundation top-level project since 2018 (daffodil.apache.org), and web search shows it has been integrated into the products of Owl Cyber Defense and Broadcom , is used in military research projects, and has a Google cloud service offering that uses it.

7.4.1.2 Clarity

a) What means are used to provide definitive descriptions beyond straight text?

The text was designed to be self-contained and includes detailed definitions and explanatory text.

b) What tables, figures and reference materials are used to remove ambiguity?

The specification is organized into tables providing core definitions for all the format properties augmented by additional sections that provide clarifying examples and discussion.

c) What contextual material is provided to educate the reader?

None is directly provided as part of the DFDL 1.0 specification.

IBM DFDL training materials are easily found on the Internet.

These materials are also available on the Internet:

- www.xfront.com tutorial
- OpenDFDL training
- User assistance with Apache Daffodil and DFDL is available via the Apache Daffodil project users mailing list.

7.4.1.3 Testability (M)

The extent, use, and availability of conformance/interoperability tests or means of implementation verification (e.g. availability of reference material for magnetic media) shall be described, as well as the provisions the specification has for testability.

The specification shall have had sufficient review over an extended time period to characterise it as being stable.

The final OGF Full Recommendation specification has been available from the OGF since February 2021. Earlier mostly-complete drafts were in use starting from September 2011.

7.4.1.4 Stability (M)

a) How long has the specification existed, unchanged, since some form of verification (e.g. prototype testing, paper analysis, full interoperability tests) has been achieved?

The Specification was ratified as a Grid Full Recommendation by the OGF in February 2021. It was a Proposed Recommendation in substantially complete form since January 2011 but took until June 2019 for two interoperable implementations to be demonstrated. The report of interoperability was published in June 2019, and the specification changes from that date until final ratification in Feb 2021 were essentially clarifications and editorial improvements, not functional content changes.

b) To what extent and for how long have products been implemented using the specification?

IBM DFDL (several different implementations) first appeared publicly in 2011. The ESA (European Space Agency) DFDL4Space first mention was 2016 (Java) and 2018 (C++). The Apache Daffodil Open-Source project began around 2009, but accelerated starting in 2012.

All these implementations were revised as the specification was refined on the way to becoming a full OGF Recommendation.

c) What mechanisms are in place to track versions, fixes and addenda?

The Specification review process occurs primarily through the Project mailing list. A formal process for creating informational documents and for proposing extensions to DFDL is also defined. The DFDL work group meets via video-conference every few weeks to discuss action items. Changes to the specification are discussed in email threads and eventually are escalated to workgroup action items. When there is a consensus on a need for a change/improvement in the DFDL specification, a GitHub tracker is opened, and remains open while specific language is proposed for addressing the matter. These issues are closed once the change has been integrated into the specification document. (With change history tracked and marked.)

Proposed draft and Released versions of the specification are made available on the OGF website, and on the DFDL GitHub site.

7.4.1.5 Availability (M)

a) Where is the specification available (e.g. one source, multinational locations, what types of distributors)?

The Specification is available at the OGF DFDL Project website: https://www.ogf.org/dfdl where the official PDF version of the specification is maintained.

Draft and ongoing work on the specification is conducted in the open on the OGF DFDL GitHub site: https://github.com/OpenGridForum/DFDL

b) How long has the specification been available?

The official OGF Full Recommendation version of the DFDL specification GFD 240 is dated February 2021. This superseded two earlier largely complete OGF Proposed Recommendation draft versions: GFD 207 was published in September 2014 and GFD 174 was published in January 2011.

c) Has the distribution been widespread or restricted? (describe the situation)

The Specification is available to all interested parties via a link on the www.ogf.org website without restriction.

d) What are the costs associated with specification availability?

The specification is available without cost.

7.4.2 Consensus (M)

The accompanying report shall describe the extent of (inter)national consensus that the document has already achieved.

7.4.2.1 Development Consensus

a) Describe the process by which the specification was developed.

Face to face meetings were held several times a year as part of OGF general meetings from roughly 2003 to 2009 at which point the spec was largely complete. Subsequently, the specification was refined and clarified primarily via regular semi-weekly telephone calls and the project mailing lists. The formal process for creating an official draft and submitting it for consideration as an OGF Recommendation is documented in OGF GFD 152.

OGF documents are developed according to an open public process as described at this link http://redmine.ogf.org/projects/editor/wiki/About_OGF_Documents . Tools to support creation of new OGF documents are described here: https://docs.ogf.org/ogf-documents-and-workspaces/ .

b) Describe the process by which the specification was approved.

This process is documented in OGF GFD 152. This includes public comment periods and multiple passes of editorial review including by an external editor/reviewer to insure public comments are addressed. Part of the review process is reviewing interoperability documentation.

c) What "levels" of approval have been obtained?

DFDL is an OGF Full Recommendation.

7.4.2.2 Response to User Requirements

a) How and when were user requirements considered and utilized?

User requirements are solicited via the regular Project calls and the mailing lists, including feedback received by the public on an ongoing basis. This feedback is directly utilized in drafting each subsequent revision of the Specification.

b) To what extent have users demonstrated satisfaction?

Approximately 1300 IBM customers are potentially using IBM DFDL for mission-critical integration & automation applications, including banks, airlines, retailers, manufacturers.

At Google, there is a cloud offering using DFDL capabilities for Application Modernization.

DFDL has been adopted for cyber-security needs by a number of vendors including Owl Cyber Defense, Broadcom and is cited in at least one cyber-security-related patent. It also plays a role in prominent military research programs though public documentation of these efforts is thin.

7.4.2.3 Market Acceptance

a) How widespread is the market acceptance today? Anticipated?

See prior items.

b) What evidence is there of market acceptance in the literature?

A search for DFDL "data format" on Google Scholar turns up a few academic papers.

Interest in DFDL has not been primarily academic, but here are two published research works that used DFDL processing as part of their research infrastructure.

- T. Strayer et al., "Mission-Centric Content Sharing Across Heterogeneous Networks," 2019 International Conference on Computing, Networking and Communications (ICNC), Honolulu, HI, USA, 2019, pp. 1034-1038, doi: 10.1109/ICCNC.2019.8685557.
- Jeter, S. et al. (2020). Semantic Links Across Distributed Heterogeneous Data. In: Herrera, F., Matsui , K., Rodríguez-González, S. (eds) Distributed Computing and Artificial Intelligence, 16th International Conference. DCAI 2019. Advances in Intelligent Systems and Computing, vol 1003 . Springer, Cham. https://doi.org/10.1007/978-3-030-23887-2_13

The public GitHub site for DFDL Schemas has two popular DFDL schemas which happen to be for ISO data formats. ISO8583 (33 stars rating) and ISO9735/EDIFACT (22 stars rating)

7.4.2.4 Credibility

a) What is the extent and use of conformance tests or means of implementation verification?

Interoperability of two implementations for 14 different DFDL schemas, including those for major data formats like ISO8583, and ISO9735/EDIFACT has been documented in OGF GWDE DFDL Experience Document 6 which was published in June 2019.

Both the IBM and Apache Daffodil implementations make use of a test language called TDML (Test Data Markup Language) which enables test to be created and run against multiple DFDL implementations.

b) What provisions does the specification have for testability?

Not applicable.

7.4.3 Alignment

The specification should be aligned with existing JTC 1 standards or ongoing work and thus complement existing standards, architectures and style guides. Any conflicts with existing standards, architectures and style guides should be made clear and justified.

7.4.3.1 Relationship to Existing Standards

a) What International Standards are closely related to the specification and how?

The DFDL 1.0 Specification is layered on top of W3C XML Schema 1.0, and the DFDL expression language is based on W3C XPath 2.0.

A related standard is ASN.1 ECN, i.e., ITU-T Recommendation X.692 | ISO/IEC 8825-3, Information technology – ASN.1 encoding rules: Specification of Encoding Control Notation (ECN). ASN1 ECN is an earlier attempt to create a declarative format description system.

b) To what International Standards is the proposed specification a natural extension?

W3C XML Schema and W3C XPath - see above.

c) How the specification is related to emerging and ongoing JTC 1 projects?

JTC 1 SC41 (Internet of Things, Digital Twin) has a clear relationship to cloud/grid computing technology which has clear relationship to the mission of the OGF generally and DFDL specifically.

7.4.3.2 Adaptability and Migration

a) What adaptations (migrations) of either the specification or International Standards would improve the relationship between the specification and International Standards?

Three ISO data format standards, EDIFACT ISO 9735, ISO 8583, and Swift ISO 15022 subset were integral to creation and acceptance of the DFDL standard at the OGF level. Going forward additional ISO data standards can be described using DFDL schemas.

b) How much flexibility does the PAS Submitter have?

The submission is made by the OGF DFDL Working Group.

c) What are the longer-range plans for new/evolving specifications?

The DFDL workgroup strives for stability in the specification, and robust compatibility between revisions.

A standard mechanism is in place for proposing extensions to DFDL. This requires development of a prototype implementation that adheres to the guidelines ensuring that the proposed extension features are clearly identifiable in the text so users do not become dependent on these features accidentally. Experience reports about real use cases and the experience with the feature are required before consideration of incorporation of an extension into a future version of the DFDL standard.

7.4.3.3 Substitution and Replacement

a) What needs exist, if any, to replace an existing International Standard? Rationale?

Not applicable.

b) What is the need and feasibility of using only a portion of the specification as an International Standard?

The DFDL 1.0 specification is not structured in a way that makes it amenable to adopting only a portion as an ISO standard.

From the perspective of an implementor using the DFDL specification, DFDL is a large standard; hence, the DFDL 1.0 specification explicitly allows *conforming* subsets, and identifies the optional features that can be omitted from an implementation while still conforming to the DFDL specification.

c) What portions, if any, of the specification do not belong in an International Standard (e.g. too implementation-specific)?

There are no sections that are too implementation specific as witnessed by the existence of multiple separate implementations.

7.4.3.4 Document Format and Style

a) What plans, if any, exist to conform to JTC 1 document styles?

The next significant revision of the DFDL specification will be created with the standard JTC 1 document style as a primary output format.

Experience creating DFDL schemas for data formats has taught us that all large specifications should be machine-readable and processable so that formal artifacts like interactive help systems, rapid contextual access to the specification text from user interfaces, and even parts of the implementations can be generated from the machine-readable form of the specification. Our goal for the next major version is to generate the JTC 1 document style, and to enable generation of other formal artifacts, from a common machine-readable form of the specification, likely specified in XML.

7.4.4 Maintenance (M)

a) Have changes occurred on the subject of maintenance since the PAS Submitter application or renewal, or for a Fast Track, since the most recent submission of the standard? (This is the place to mention any particular agreement reached with a JTC 1 subgroup).

No.

END