FAIR DIGITAL OBJECTS FORUM
https://fairdo.org/

FDO Machine Actionability

Version 2.0

FDO Forum Proposed Recommendation 11 June 2022

Current version:
https://docs.google.com/document/d/1GHFPAUGpNvYaxctkx-CpvY1vKf_aGZpSlWGOWvyXSiQ/edit

Previous version:
https://docs.google.com/document/d/1d7rF6IERtrky8D2nTFFUfAPJxWWRQtiLrksG4FIC2yA/edit

Editors:

Claus Weiland, Peter Wittenburg

Authors:

Claus Weiland, Sharif Islam, Daan Broder, Ivonne Anders, Peter Wittenburg

Abstract

The FDO Forum will refer in many of its documents to the term “machine actionability” which was coined in an article by Wittenburg, Strawn, Mons, Boninho and Schultes[1] and later exemplified by GO FAIR implementation networks in the realm of the FAIR principles. For the FDO work we need a clear definition of what is meant with the term “machine actionability”. This document includes a verbose description of relevant terms and draws conclusions in the form of a definition for use in the FDO Forum.

[1] http://doi.org/10.23728/b2share.b605d85809ca45679b110719b6c6cb11

Machine actionability, in the context of FDO, represents the FDO Forum community's focus on architecting the FDO to maximize dynamic data accessibility, parsability, interpretability, understandability, and synthesizability by machines while minimizing or eliminating the need for human oversight.

Status of this document

This document is a Working Draft of a recommendation track document that has been discussed in the FDO TSIG Working Group as WD 0.2 and therefore needs to be discussed and endorsed now by the FDO Forum as version WD 1.2.

We will start using the FDO Share to host the documents in the different phases.

https://datashare.rzg.mpg.de/apps/files/?dir=/FDO-Forum/FDO-Document-Store&fileid=206078478

For the FDO Forum internal reviews we will use the following Google folder:
https://datashare.rzg.mpg.de/apps/files/?dir=/FDO-Forum/FDO-Document-Store/FDO%20WD-1.x&fileid=206079098

Table of contents

Abstract 1

Status of this document 2

Acknowledgments 2

1. Conclusions 3

2. Generics 4

2.1 Usage of Terms 4

2.2 Distinguishing Terms 5

3. Machine Actionability Definitions 6

4. FDO and Machine Actionability 7

5. Changes from previous versions 9

1. Conclusions

In this chapter we will draw a set of conclusions from the descriptions below.

1. There are yet no clear conceptual distinctions and widely agreed definitions about the terms “machine readability”, “machine interpretability” and “machine actionability”.

2. For FDO Forum purposes we agree on the following definitions:

    a. “machine readable” are those elements in bit-sequences that are clearly defined by structural specifications.

    b. “machine interpretable” are those elements that are machine readable and can be related with semantic artifacts in a given context[2] and therefore have a defined purpose[3].

    c. “machine actionable” are those elements in bit-sequences that are machine interpretable and belong to a type for which operations have been specified in symbolic grammar[4].

3. We need to acknowledge that for FDOs there are organizational entities with PID records that have [Kernel Attributes](https://doi.org/10.15497/rda00031)[5]. After PID resolution, these attributes provide metadata information -- either directly or by pointing to metadata descriptions that need interpretation. Therefore, the FDO machine actionability in the sense of FAIRness requires 3 levels of checks:

    a. Does the resolution of a PID happen according to a standardized protocol and is the resolution result predictable and appropriate according to a specified profile[6]?

    b. Does the resulting resolution entail machine actionable attributes, i.e., can the attributes be interpreted and lead to actions based on clear definitions and typing? Are the results obtained machine actionable, i.e., can the attributes be interpreted and lead to actions based on clear definitions and typing? This must include references to essential metadata descriptions (descriptive, scientific, rights/licenses, access permissions, etc.).

---

[2] E.g. a named graph
[3] It should be noted that this also includes inferences for the identified elements.
[4] It should be noted that this also includes qualified mappings and thus derived types.
[5] https://doi.org/10.15497/rda00031
[6] See FDOF Document on PIDProfileAttributes
https://docs.google.com/document/d/1c2mZziq5pIPmLxMHLcYqlWrjYsc2ezGMXvp0E46iljo/edit#

c. Are the metadata descriptions accessible and FAIR compliant[7], i.e., that in the full sense of FAIR all specifications[8] are machine actionable if they are embedded in an FDO infrastructure that implements policies, rules and procedures for FAIR[9]?

It should be noted that 3c will be difficult to turn into practice since metadata will be created and managed by different communities and it will not be simple to change community practices. Therefore, the FDO Forum needs to find mechanisms to bridge between requirements and practices.

2. Generics

2.1 Usage of Terms

The term "machine readability" has been defined by some organizations for some time[10] while the term "machine actionability" has been introduced recently in the realm of the FAIR principles[11] and "machine-actionable data management plans" (maDMP)[12]. Some communities are also using the term "machine interpretability"[13] which implies a change of the meaning of the term "machine readability" as defined in Wikipedia in so far as the latter is then only used to indicate the capacity to transform the sequence of bits into symbols, numbers and delimiters. "Machine interpretability" will then go one step beyond this reduced definition of "machine readable" and require the FAIRness of metadata for example, i.e., the used metadata categories need to be defined and registered.

At Wikipedia we can find the following definition: ***machine-readable data****, or* ***computer-readable data****, is data in a format that can be processed by a computer. Machine-readable data must be structured data. In the United States, the OPEN Government Data Act of 14 January 2019*[14] *defines machine-readable data as "data in a format that can be easily processed by a computer without human intervention while ensuring no semantic*

---

[7] FAIR compliance can be considered from multiple domains and perspectives. There's no globally acceptable FAIR compliant tool. See https://doi.org/10.3233/ISU-200083 https://doi.org/10.1016/j.patter.2021.100370 https://doi.org/10.1038/s41597-019-0184-5

[8] This implies F2: providing rich metadata, I3: qualified cross-references, and R1: relevant attributes.

[9] Cp. Larry Lannom *State of FDO Work* SciDataCon 18 Oct 2021 https://vimeo.com/636516496

https://drive.google.com/drive/folders/1A4vbovXI0h-x_B66XfI12TzmXaB2Lay1

[10] We can refer to the formulation found in Wikipedia (https://en.wikipedia.org/wiki/Machine-readable_data)

[11] https://doi.org/10.1038/sdata.2016.18

[12] In this document we will not elaborate maDMP further; for an overview, see https://doi.org/10.1371/journal.pcbi.1006750

[13] This term is used for example in the report "Turning FAIR into Reality" (page 40, https://doi.org/10.2777/1524) and should not be mixed with the term "interpretability" which is heavily used in machine learning to make statements about trained models.

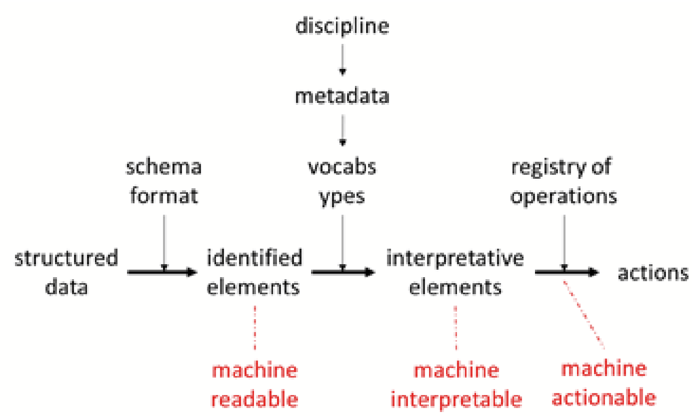
[14] https://www.cio.gov/handbook/it-laws/ogda/

meaning is lost." The law directs U.S. federal agencies to publish public data in such a manner, ensuring that "any public data asset of the agency is machine-readable.

A. Hardisty stated the following[15]: ***Machine-readable form*** *(such as RDF/XML or JSON-LD) is linking data elements to terms in vocabularies and concepts in ontologies. The form tells a machine the meaning of a data or metadata element e.g., it's a temperature, it's a license. It makes something capable of being processed; understand it; grasp the meaning of it.* ***BUT*** *it doesn't tell a machine what to do or how to behave in respect of that data element.*

Recently, the term "machine-actionable" is often used in description to indicate FAIR compliance[16]. From these usages and inspired by the discussions about the FAIR principles, we can conclude that these three terms ("machine readability", "machine interpretability" and "machine actionability") are not well defined and used in different ways by different communities. The differences in usage of these terms requires that we will first discuss what is meant with "machine readability", "machine interpretability" and then discuss "machine actionability".

2.2 Distinguishing Terms



This figure indicates the sequence of steps from identifying elements in sequences of bits making elements readable, to associating elements with defined and registered terms in semantic artefacts making the elements interpretable up to linking types of elements with operations which makes them machine actionable.

Digital Data is available in structured or unstructured forms. In unstructured documents, we find sequences of characters (text) encoded according to some character encoding system (currently mostly a Unicode specification such as UTF-8). In structured documents, we can find text fragments or numbers according to some type[17] specification (integer, float, etc.). Also, in structured documents the schema and/or format specifies how the bit sequences can be read and turned into identified structural elements, i.e., it specifies the encoding, delimiters and structure. This information is sufficient to identify elements which can be textual tokens, numbers, etc. "Machine readability" could indicate that all elements in a bit-sequence could be identified.

---

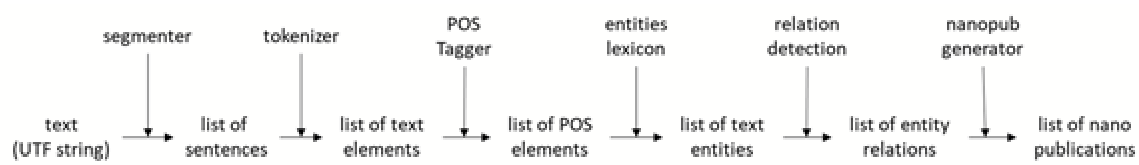
[15] Internal communications

[16] Jeffery, et.al.: Not Ready for Convergence in Data Infrastructures. Data Intelligence (2021) 3 (1): 116–135, https://doi.org/10.1162/dint_a_00084

[17] "Type" can be thought of as "characterization of data structure, contexts, and assumptions" (see Larry Lannom; Daan Broeder; Giridhar Manepalli (2015): Data Type Registries working group output. https://doi.org/10.15497/A5BCD108-ECC4-41BE-91A7-20112FF77458

Such elements can be "interpreted" if their meaning has been specified in some semantic artifact (corresponding to a *schema* or framework like the logical structure of a database, table specification, ontology, etc.). Such specifications are usually discipline/community specific, and metadata associated with data will inform users where to find these semantic artifacts. The result is what we can call "interpretative elements", i.e., humans know what is meant. It could be, for example, a text unit referring to a person or a number in a table encoding a temperature. In general, we can say that the machine identified some types based on human specifications which include contextual information (metadata)[18]. If we as humans have created a registry of operations that relates such types with actions, then machines can turn interpretative elements into actions, thus we can speak about "machine actionability".

For unstructured data like texts, Natural Language Processing (NLP) can be a guide for capturing context and meaning. A typical NLP processing sequence to generate, for example, nanopublications[19] (augmented RDF assertions) from sentences is indicated in the diagram below. At different steps of such processing, we can speak of machine readable data. For instance, when the POS (part-of-speech) tagged elements have been matched with lexical entries or a list of named entities, the identified units are associated with semantics. It would allow a machine to detect semantic relationships based on principles such as term or graph matching. And thus to create nanopublications which are formal representations of relationships augmented with administrative data.



This figure gives an impression of the typical processing steps needed to come from a sequence of symbols in unstructured documents to nano-publications, for example, which are condensed and formal descriptions of essential relationships between core terms found.

From the above example, it is obvious that the term "machine-readable" is associated with the state of "identifiable elements" as indicated in the first diagram for structured data. For textual data, the term "machine-readable" has a different meaning as the knowledge embedded in all NLP steps makes the elements at each step machine readable. This also holds for the resulting nanopublications which are in syntactical form and with semantics that humans can interpret and machines can use, for example, to draw relationships with or compute statistics on other nanopublications encoding semantics

[18] Such interpretive elements, once identified are then amenable to curation by scientific research infrastructures (see:https://doi.org/10.5334/dsj-2018-021)
[19] Nanopublications are a format for small data publications based on RDF triples, consisting of three parts: the assertion, provenance, and publication information (see: http://nanopub.org/wordpress/?page_id=65 andhttps://doi.org/10.7717/peerj-cs.78)

It may be of interest to refer to the ongoing discussion about the question whether machines actually do semantic processing or as L. Floridi argues in his book *The Fourth Revolution*[20] that our current computers are basically syntactic engines. As humans we can use machines to store semantic content, but finally it's humans who need to interpret that content. This implies that when a relation to a semantic artifact indicates that a value, for example, is a temperature measured in Celsius at a certain place, height, and context, we as humans can interpret the value. Machines are able to validate and consequently process these values without understanding -in the intellectual sense- what they are exactly about.

3. Machine Actionability Definitions

The FAIR principles require further elaboration of the term "machine actionability", another step beyond the steps described above, independent of how the terms "machine readable" or "machine interpretable" are used. Using search engines and a document from DiSSCo we find three statements:

The DDI Initiative states[21]: *This term **machine actionable** refers to information that is structured consistently so that machines, or computers, can be programmed against the structure. DDI provides machine-actionable metadata.* It is obvious that this definition does not go beyond what is stated above as machine readable. Therefore, this statement is not sufficient.

The GO FAIR website[22] states: *The FAIR principles emphasize **machine-actionability** (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.* This definition purposely provides a high level description of machine-actionability and does not refer to concrete implementation options, i.e., it does not speak explicitly about linking with operations.

A. Hardisty stated the following[23]: *The ability of a machine to know what actions to perform and how to behave in relation to some data element(s). (1) Encoding of something that gives cause for action. (2) Tells a machine what subset of decisions to apply, taken from a set of possible alternatives*. This definition indicates what needs to be done in addition to allowing machines to carry out some operations using the interpretative element. So it broadens the perspective of the GO FAIR statement towards Or if it finds a "temperature measured at 2 m above sea level" , transform it to a type "temperature at 10 m above sea level" according to some algorithm using additional data. As Hardisty indicates, to make the step from something being **machine interpretable** to something being **machine actionable** there needs to be a link between the type of the attributes used in a structure with an operation that can act on the values found i.e.,

[20] Floridi, L. (2014). *The 4th revolution: How the infosphere is reshaping human reality*. https://identifiers.org/isbn:9780199606726
[21] https://ddialliance.org/taxonomy/term/198
[22] https://www.go-fair.org/fair-principles
[23] Internal communication

be invoked and achieve some result as a consequence. In short: machine actionability refers to the knowledge about "how" to further proceed.

In general cases, rich contextual and provenance information will be required to interpret a found value "correctly". The term "correctly" includes a variety of aspects (creators intention, contextual biases, licensing information determining usage restrictions, etc.) which we will not elaborate on in this document. It should be noted that users could want to pick different possible operators working on the same set of values such as different machine learning packages.

4. FDO and Machine Actionability

Here we look only at the DO based approach of the FDO model, similar statements could be made for the LD approach to FDOs. As the well-known diagram indicates, FDOs are organizations around bit sequences that bind all kinds of relevant information about this bit sequence together. It should be noted that FDOs are absolutely agnostic to whatever the bit-sequences are encoding and how they are encoding content. However, papers such as FDO Configuration Types[24] indicate that there are different possible ways to configure FDOs and nevertheless fulfill the FDO Framework specifications[25]. Therefore, when applying the term "machine actionability" to FDOs we need to check for all steps in disclosing the rich information included in an FDO what "machine actionability" actually means. This is stated in the FDO Framework specifications and is also clearly explained in presentations given by L. Bonino[26], for example.

We can distinguish 3 levels:

1. A software client (machine) finds a PID and needs to be told what to do with it. It will be sent to a resolver which sends back a set of key-value pairs such as "cksm = xxxxxxxxxxxxx" or "metadata = http://xyz.uvw/mymetadata".
2. The software client now needs to have a mechanism which enables taking actions on a value that is associated with the label "cksm", for example. If the term "cksm" is defined and registered as being a checksum of type MD5, the machine could link this with an operation to check authenticity. The necessary link can be embedded in the client software or the client software will know to look in a specific registry for operations for this type. The software client also needs to have a mechanism which enables taking actions using the link specified as a value of the type metadata. It will understand that

[24] See FDOF Working Document on ConfigurationTypes https://docs.google.com/document/d/1ivvg3C_QWSO9PIQwkKT89xG4fBhNAs7_6b0Dz11EwDg/edit

[25] https://github.com/GEDE-RDA-Europe/GEDE/blob/master/FAIR%20Digital%20Objects/FDOF/FAIR%20Digital%20Object%20Framework-v1-02.docx

[26] https://github.com/GEDE-RDA-Europe/GEDE/blob/master/FAIR%20Digital%20Objects/Paris-FDO-workshop/GEDE_Paris_Session%201_Bonino.pptx

this is an identifier which can be resolved to receive the metadata description. ~~This is done using a http-get command or a PID if the metadata is organized as a self-standing FDO.~~

3. The software client receives the metadata description created according to some community schema standard. The corresponding metadata categories should be properly defined and registered in an open registry, i.e. the metadata assertions are machine readable and machine interpretable.To meet the requirements of FAIR Digital Objects also the metadata also needs to be machine actionable. For example, they may contain the specifications of where to find the reference to the object's bit-sequence, the license statement, the record with authorisation information, etc. If there is no clear type definition indicating where to find this information and if there is no relation between the type and an operation, we can hardly speak about a high level of FAIR compliance.

From the definition of the term "machine actionable" as indicated by Hardisty and from the nature of FDOs it is obvious that in principle we can speak about 3 levels of FAIR compliance as indicated above. Achieving the first 2 levels is implicitly given when implementing FDOs properly. Achieving full compliance is dependent on the existence of standardized and operationally-used discipline-specific metadata in the various scientific communities.

T5. Changes from previous versions

0.1	TSIG members	2021-10-05	The draft emerged from an interaction between Daan Broder, George Strawn and Peter Wittenburg based on documents from amongst others A. Hardisty & L. Lannom. All TSIG members contributed then to version 0.1.
0.2	TSIG-Cochairs	2021-10-11	TSIG co-chairs added conclusions

1.1	FDO Forum	2022-02-25	FDO F internal review
2.0	TSIG	2022-06-11	Iterative discussion in TSIG