Metadata Standards & Digital Object Types

FAIR Digtial Objects are self-describing Digital Objects



 FAIR Implementation Strategies

 GO FAIR Foundation

 http://orcid.org/0000-0001-8888-635X

 erik.schultes@go-fair.org

 go-fair.org

Dr. Erik Schultes, PhD

LCDS LEIDEN CENTRE OF DATA SCIENCE

April 29, 2021 Slides: https://osf.io/tp25n/ Inclusion of METADATA in general ...

We noticed that many repositories do not provide enough metadata to describe the data and we also see that the object types are not properly indicated - is it a graph, excel, publication, report or something else. Inclusion of METADATA in general ...

We noticed that many repositories do not provide enough metadata to describe the data and we also see that the object types are not properly indicated - is it a graph, excel, publication, report or something else.

Ipload type								required 💙
Publication	Doster	Presentation		Image	Uideo/Audio	>Software	Lesson	Physical object
o	0	0	0	े 🗰 Other	0	0	0	0
				0				
I	Publication ty	Journal ar	ticle					\$
		Publication	type.					

Box 2 | The FAIR Guiding Principles

To be Findable:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles

13. (meta)data include qualified references to other (meta)data

To be Reusable:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

FAIR Data Point

https://www.fairdatapoint.org

About FAIR Data Point

FAIR Data Point

What is a FAIR Data Point

A FAIR Data Point (sometimes abbreviated to FDP) is the realisation of the vision of a group of authors of the original paper on FAIR on how (meta)data could be presented on the web using existing standards, and without the need of APIs.

A FAIR Data Point ultimately stores *information abou* And just like the *webserver* in the WWW in the begin publishing text to anyone, a FAIR data point aims to g on the web.

The system is called a **FAIR** data point because it tak taken care of to make data FAIR; especially with the and a uniform open way of **A**ccessing the data. The I of the metadata it stores, but it leaves the Interopera provider.

Components

Search for great content (e.g., mysql)
 Explore fairdata/fairdatapoint

https://hub.docker.com/r/fairdata/fairdatapoint



The FAIR Data Point as we have implemented it has three components.







Title / Description / Publisher / Version / Language / License / Start date / Last update / Institution / Country



Data Catalog layer

Title / Description / Publisher / Version / Language

Dataset layer

Title / Description / Publisher / Version / Language / License / Issued / Modified / Keywords / Theme / Contact point / Landing page

Distribution layer

Title / Description / License / Issued / Modified / Media Type / Format / Byte Size / Download URL / Access URL

https://github.com/FAIRDataTeam/FAIRDataPoint-Spec http://www.w3.org/TR/vocab-dcat/

Title / Description / Publisher / Version / Language / License / Start date / Last update / Institution / Country



Data Catalog layer

Title / Description / Publisher / Version / Language

Dataset layer

Title / Description / Publisher / Version / Language / License / Issued / Modified / Keywords / Theme / Contact point / Landing page

Distribution layer

Title / Description / License / Issued / Modified / Media Type / Format / Byte Size / Download URL / Access URL

https://github.com/FAIRDataTeam/FAIRDataPoint-Spec http://www.w3.org/TR/vocab-dcat/



FAIR Data Point A1.1, A1.2

Title / Description / Publisher / Version / Language / License / Start date / Last update / Institution / Country



Data Catalog layer

Title / Description / Publisher / Version / Language

Dataset layer

Title / Description / Publisher / Version / Language / License / Issued / Modified / Keywords / Theme / Contact point / Landing page

Distribution layer

Title / Description / License / Issued / Modified / Media Type / Format / Byte Size / Download URL / Access URL

Title / Description / Publisher / Version / Language / License / Start date / Last update / Institution / Country



Data Catalog layer

Title / Description / Publisher / Version / Language

R1.1

Dataset layer

Title / Description / Publisher / Version / Language / License / Issued / Modified / Keywords / Theme / Contact point / Landing page

Distribution layer

Title / Description / License / Issued / Modified / Media Type / Format / Byte Size / Download URL / Access URL

Title / Description / Publisher / Version / Language / License / Start date / Last update / Institution / Country



Data Catalog layer

R1.2

Title / Description / Publisher / Version / Language

Dataset layer

Title / Description / Publisher / Version / Language / License / Issued / Modified / Keywords / Theme / Contact point / Landing page

Distribution layer

Title / Description / License / Issued / Modified / Media Type / Format / Byte Size / Download URL / Access URL

Title / Description / Publisher / Version / Language / License / Start date / Last update / Institution / Country



Data Catalog layer

Title / Description / Publisher / Version / Language

Dataset layer

Title / Description / Publisher / Version / Language / License / Issued / Modified / Keywords / Theme / Contact point / Landing page

Distribution layer

Title / Description / License / Issued / Modified / Media Type / Format / Byte Size / Download URL / Access URL



Title / Description / Publisher / Version / Language / License / Start date / Last update / Institution / Country



Data Catalog layer

Title / Description / Publisher / Version / Language

F2

Dataset layer

Title / Description / Publisher / Version / Language / License / Issued / Modified / Keywords / Theme / Contact point / Landing page

Distribution layer

Title / Description / License / Issued / Modified / Media Type / Format / Byte Size / Download URL / Access URL

Title / Description / Publisher / Version / Language / License / Start date / Last update / Institution / Country



Data Catalog layer

Title / Description / Publisher / Version / Language

F2

Dataset layer

Title / Description / Publisher / Version / Language / License / Issued / Modified / Keywords / Theme / Contact point / Landing page

Distribution layer

Title / Description / License / Issued / Modified / Media Type / Format / Byte Size / Download URL / Access URL

graph, excel, publication, report

F/IR FAIR Data Point

KIU FAIR Data Point

This FAIR Data Point contains the metadata of datasets and other artefacts related to the COVID-19 pandemic and other health data in Kampala International University and the rest of Uganda.

Catalogs

Covid-19 Case Report Form

Covid-19 case report forms following WHO standard

Issued 13-07-2020 Modified 13-07-2020

Migrant Media Reports

These are national and international media reports published during the covid-19 crisis.

Issued 13-07-2020 Modified 13-07-2020

Metadata IssuedMetadata Modified01-07-202013-07-2020

Version

Language

<u>en</u>

1.0

License

<u>cc-by-nc-nd3.0</u>

Institution <u>kiu.ac.ug</u>

Start date 03-06-2020

Last update 13-07-2020

Institution country

<u>Q1036</u>



This FAIR Data Point contains the metadata of datasets and other artefacts related to the COVID-19 pandemic and other health data in Kampala International University and the rest of Uganda.

Catalogs

Covid-19 Case Report Form

Covid-19 case report forms following WHO standard

Issued 13-07-2020 Modified 13-07-2020

Migrant Media Reports

These are national and international media reports published during the covid-19 crisis.

Issued 13-07-2020 Modified 13-07-2020

Metadata Issued Metadata Modified 01-07-2020 13-07-2020

Version

1.0

Language

<u>en</u>

License <u>cc-by-nc-nd3.0</u>

Institution <u>kiu.ac.ug</u>

Start date 03-06-2020

Last update 13-07-2020

Institution country

<u>Q1036</u>

F/**IR** FAIR Data Point



KIU FAIR Data Point / Migrant Media Reports

Migrant Media Reports

These are national and international media reports published during the covid-19 crisis.

Datasets

Media Report Dataset

These datasets are media reports concerning events related to the impact of covid-19 crisis on immigrants in Tunisia.

Q38926 Q57979909

Issued 13-07-2020 **Modified** 13-07-2020

Metadata Issued 13-07-2020	Metadata Modified 13-07-2020
Version 1.0	
Language <u>en</u>	

License

Issued 13-07-2020

Modified

13-07-2020

Theme taxonomy

- <u>Q38926</u>
- <u>Q57979909</u>

F/IR FAIR Data Point



KIU FAIR Data Point / Migrant Media Reports / Media Report Dataset

Media Report Dataset

These datasets are media reports concerning events related to the impact of covid-19 crisis on immigrants in Tunisia.

Distributions

Media Reports listed in a google document

These are internet links to media reports listed in a google document.

Issued 13-07-2020 Modified 13-07-2020 Media Type https://w3id.org/mediatype/text/html

Metadata Modified 13-07-2020

F/IR FAIR Data Point



Issued 13-07-2020 Log in

Modified

13-07-2020

Theme

- <u>Q38926</u>
- <u>Q57979909</u>

F/IR FAIR Data Point

Access KIU FAIR Data Point / Migrant Media Reports / Media Report Dataset / Media Reports listed in a google do... to the Media Reports listed in a google document dataset These are internet links to media reports listed in a google document. 🗹 Access online Metadata Issued Metadata Modified 13-07-2020 13-07-2020 Version 1.0 Language <u>en</u> License Issued 13-07-2020 Modified 13-07-2020 Media type https://w3id.org/mediatype/text/html



6.7 Class: Distribution

The following properties are specific to this class: access rights, access URL, access service, byte size, compression format, conforms to, description, download URL, format, has policy, license, media type, packaging format, release date, rights, spatial resolution, temporal resolution, title, update/modification date.

RDF class:	dcat:Distribution
Definition:	A specific representation of a dataset. A dataset might be available in multiple serializations that may differ in various ways, including natural language, media-type or format, schematic organization, temporal and spatial resolution, level of detail or profiles (which might specify any or all of the above).
Usage note:	This represents a general availability of a dataset. It implies no information about the actual access method of the data, i.e. whether by direct download, API, or through a Web page. The use of dcat:downloadURL property indicates directly downloadable distributions.
See also:	§ 6.8 Class: Data Service

NOTE

Examples of distributions include a CSV file, a [netCDF] file, a JSON document, or a data-cube, files made accessible according to different profiles, such as XML or JSON schemas or [ShEx] or [SHACL] expressions.

6.7 Class: Distribution

The following properties are specific to this class: access rights, access URL, access service, byte size, compression format, conforms to, description, download URL, format, has policy, license, media type, packaging format, release date, rights, spatial resolution, temporal resolution, title, update/modification date.

RDF class:	dcat:Distribution
Definition:	A specific representation of a dataset. A dataset might be available in multiple serializations that may differ in various ways, including natural language, media-type or format, schematic organization, temporal and spatial resolution, level of detail or profiles (which might specify any or all of the above).
Usage note:	This represents a general availability of a dataset. It implies no information about the actual access method of the data, i.e. whether by direct download, API, or through a Web page. The use of <u>dcat:downloadURL</u> property indicates directly downloadable distributions.
See also:	§ 6.8 Class: Data Service

NOTE

Examples of distributions include a CSV file, a [netCDF] file, a JSON document, or a data-cube, files made accessible according to different profiles, such as XML or JSON schemas or [ShEx] or [SHACL] expressions.

Dublin Core[™]Metadata Initiative

https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#http://purl.org/dc/terms/format

Term Name: format More details			
URI	http://purl.org/dc/terms/format		
Label	Format		
Definition	The file format, physical medium, or dimensions of the resource.		
Comment	Recommended practice is to use a controlled vocabulary where available. For example, for file formats one could use the list of Internet Media Types [MIME]. Examples of dimensions include size and duration.		

Media Types

Last Updated

2021-04-28 Registration Procedure(s)

Expert Review for Vendor and Personal Trees.

Expert(s)

Ned Freed, Alexey Melnikov, Murray Kucherawy (backup)

Reference

[RFC6838][RFC4855]

Note

Per Section 3.1 of [RFC6838], Standards Tree requests made through IETF documents will be reviewed and approved by the IESG, while requests made by other recognized standards organizations will be reviewed by the Designated Expert in accordance with the Specification Required policy. IANA will verify that this organization is recognized as a standards organization by the IESG.

Note

[RFC2046] specifies that Media Types (formerly known as MIME types) and Media Subtypes will be assigned and listed by the IANA.

Procedures for registering Media Types can be found in [<u>RFC6838</u>], [<u>RFC4289</u>], and [<u>RFC6657</u>]. Additional procedures for registering media types for transfer via Real-time Transport Protocol (RTP) can be found in [<u>RFC4855</u>].

The following is the list of Directories of Content Types and Subtypes. If you wish to register a Media Type with the IANA, please see the following for the online application:

[Application for registration of Media Types]

Other Media Type Parameters: [<u>IANA registry media-types-parameters</u>] Media Type Sub-Parameters: [<u>IANA registry media-type-sub-parameters</u>] Provisional Standard Media Type Registry: [<u>IANA registry provisional-standard-media-types</u>]

Name

1d-interleaved-parityfec 3gpdash-qoe-report+xml 3gppHal+json 3gppHalForms+json 3gpp-ims+xml A2L activemessage activity+json alto-costmap+json alto-costmapfilter+json alto-directory+json alto-endpointprop+json alto-endpointpropparams+json alto-endpointcost+json alto-endpointcostparams+json alto-error+json alto-networkmapfilter+json alto-networkmap+json alto-updatestreamcontrol+json alto-updatestreamparams+json AML andrew-inset applefile ATF ATFX atom+xml atomcat+xml atomdeleted+xml atomicmail

atomsvc+xml

Template application/1d-interleaved-parityfec application/3gpdash-goe-report+xml application/3gppHal+json application/3gppHalForms+json application/3gpp-ims+xml application/A2L application/activemessage application/activity+json application/alto-costmap+json application/alto-costmapfilter+json application/alto-directory+json application/alto-endpointprop+json application/alto-endpointpropparams+json application/alto-endpointcost+json application/alto-endpointcostparams+json application/alto-error+json application/alto-networkmapfilter+json application/alto-networkmap+json application/alto-updatestreamcontrol+json application/alto-updatestreamparams+json application/AML application/andrew-inset application/applefile application/ATF application/ATFX application/atom+xml application/atomcat+xml application/atomdeleted+xml application/atomicmail application/atomsvc+xml

Reference

[RFC6015] [_3GPP][Ozgur_Oyman] [_3GPP][Ulrich_Wiehe] [3GPP][Ulrich Wiehe] [_3GPP][John_M_Meredith] [ASAM][Thomas_Thomsen] [Ehud Shapiro] [W3C][Benjamin_Goering] [RFC7285] [RFC7285] [RFC7285] [RFC7285] [RFC7285] [RFC7285] [RFC7285] [RFC7285] **RFC7285 RFC7285** [RFC8895] [RFC8895] [ASAM][Thomas_Thomsen] [Nathaniel Borenstein] [Patrik Faltstrom] [ASAM][Thomas_Thomsen] [ASAM] [Thomas_Thomsen] [RFC4287][RFC5023] [RFC5023] [RFC6721] [Nathaniel Borenstein] [RFC5023]



Metadata for Machines Workshops https://www.go-fair.org/today/making-fair-metadata/





Metadata for Machines Workshops https://www.go-fair.org/today/making-fair-metadata/





Metadata for Machines Workshops https://www.go-fair.org/today/making-fair-metadata/

Data Stream **Data Source** Resource Type _ Variable Dataset Identifier Spatial Coverage **Related Resource Vertical Coverage** Version **Temporal Coverage** Dataset Language Metadata Title 'Extension' elements Description Subjects and Keywords Dataset Creator Distribution Contributor Metadata Publisher Distribution Rights Date Funding 'Dataset distribution' element(s) 'Typical' elements

Elements that we will re-create to turn GDMT into a domain specific template



Nikola Vasiljevic - Introduction to Metadata

R1.3 & Certification





FAIR well !

Acknowledgements

CODEVENCE



Barbara Magagna Environment Agency Austria (PhD student Univ. Twente)



Nikola Vasiljevic FAIR Data Collective

Jeroen Belien, AUMC John Graybeal, Stanford Univeristy Kristina Hettne, Leiden University Tobias Kuhn, Assistant Professor at Vrije Universiteit Amsterdam Robert Pergl, Czech Technical University in Prague Jan Slifka, Czech Technical University in Prague Wouter van Stralen, Low Code Solutions Tieneke Schaaij-Visser, Lygature, Health-RI Sofie Hansen, Lygature, Health-RI Rita Azevedo, Lygature, Health-RI Mijke Jetten, DTL, Health-RI International Science Coordinator

> VRIJE UNIVERSITEIT AMSTERDAM

DTL >>>



DeiC



Certification Process



Manobench I my channel | others | search | publish |

Publish a new Nanopublication

Assertion: Certifying a nanopublication <a> (change)

https://w3i	d.org/fair/gofair/terms/GO-FAIR-Foundation	certifies	http://purl.org/np/RAAEKgxLdgvZ3q9bsg0-Fcf8s86jH92YGWYOtNJn6nyKU		
Provenance:	Attributed to myself	▼ 7			
The assertion above is attributed to me .					
Publication in	add element	Ŧ			
This nanopublication is created by me .					
I understand that publishing cannot be undone and that the provided information will be publicly visible and openly connected to my ORCID identifier.					
Publish					

Approves-of

• Certifies (conforms to GFF requirements)