Building & Mining Knowledge Graphs

(KEN4256)

Lecture 6: Publishing Knowledge Graphs

Maastricht University
Institute of Data Science

id: KEN4256_L6
version: 1.2024.0
created: January 13, 2019
last modified: March 26, 2024
published on: March 26, 2024

Significant **effort is needed to <u>find</u> the right information, make <u>sense</u> of it, <u>understand</u> your rights and responsibilities, and ultimately <u>reuse</u> them for a new purpose**

https://youtu.be/N2zK3sAtr-4

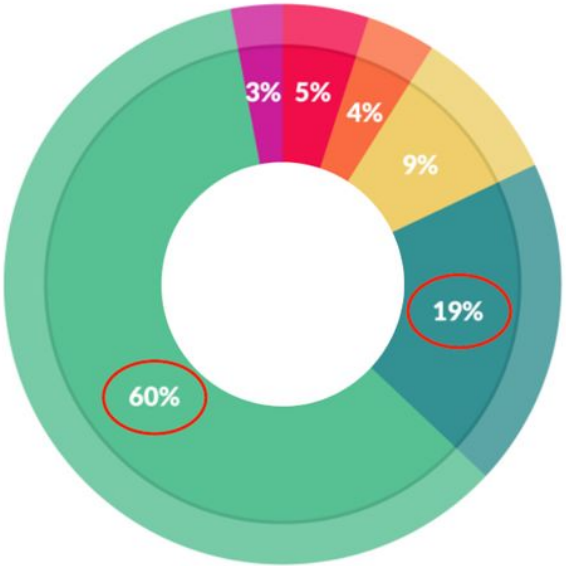
slido

What problems were encountered

Click **Present with Slido** or install our Chrome extension to activate this poll while presenting.

What problems were there?

- Limited data availability
- Inefficient data access
  - Unreliable (non-replicated) storage
- Proprietary data format
- Non-standardized metadata
  - No description of the data elements
  - Lack of contact information
  - No traceability of data from publication to source

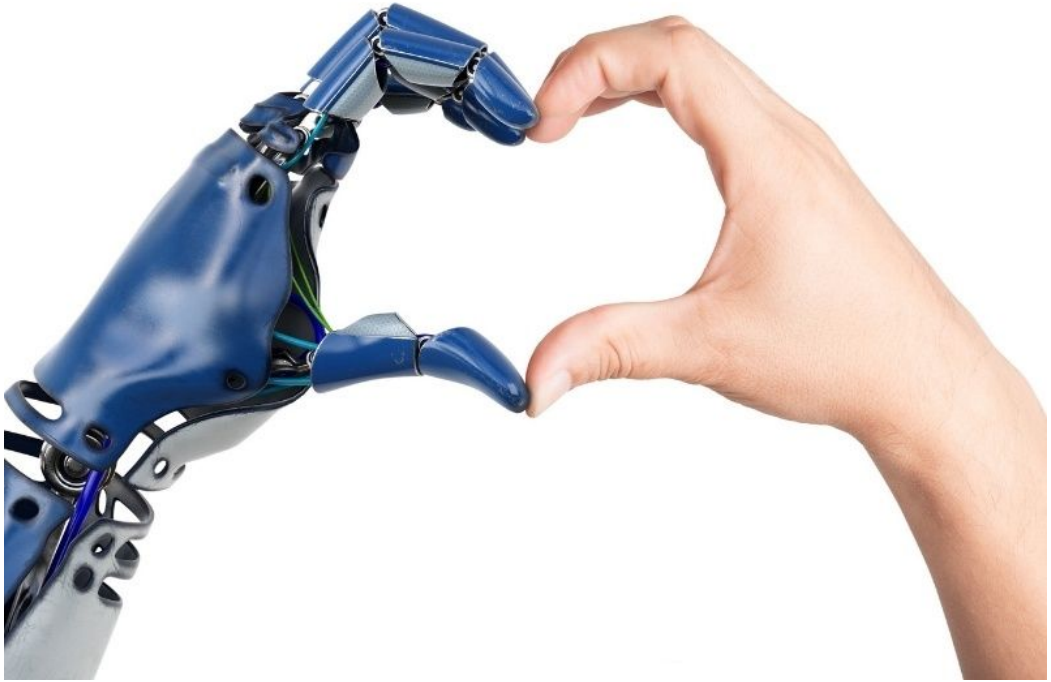


What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

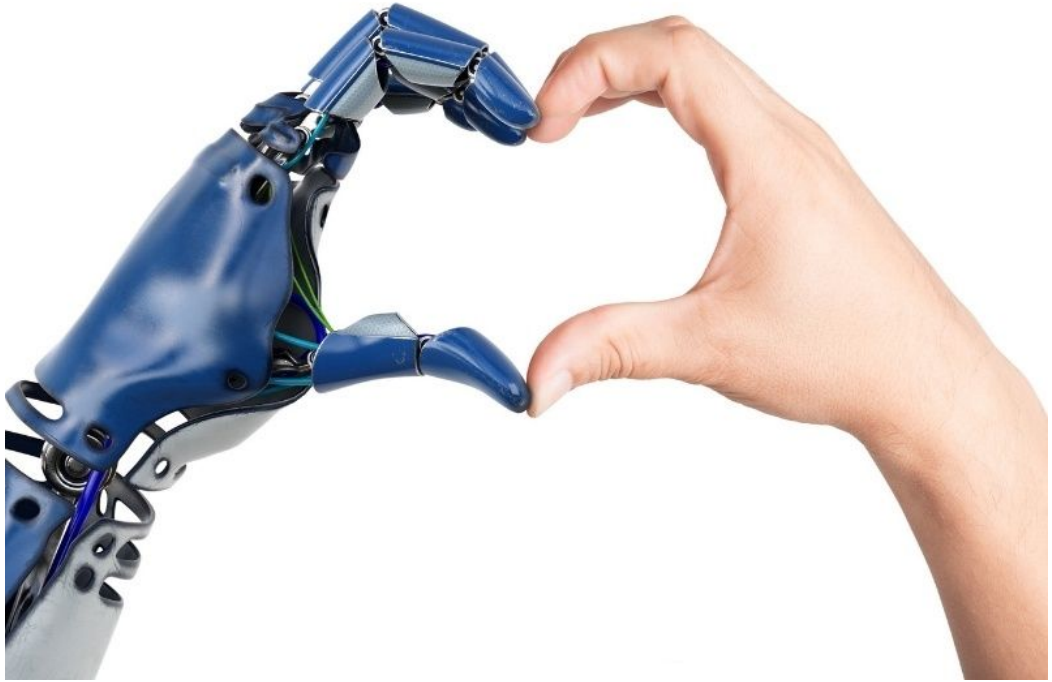
http://visit.crowdflower.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf

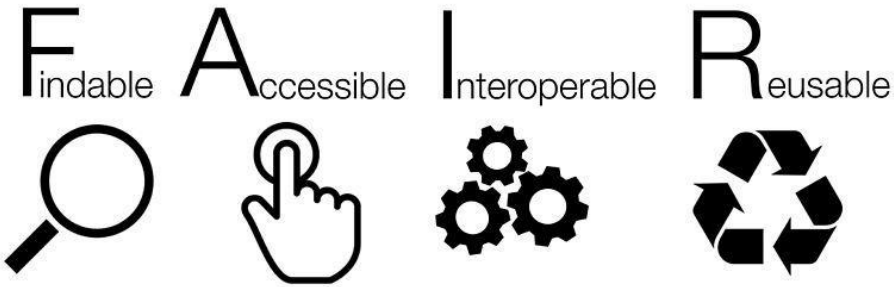
Human Machine collaboration is crucial to our future work



Machines

need to be able to discover and reuse data (and arguably any digital resource)





An international, bottom-up <u>paradigm</u> for the discovery and reuse of digital content

by and for people and machines

SCIENTIFIC DATA

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier [...] Barend Mons

Affiliations | Contributions | Corresponding author

Scientific Data **3**, Article number: 160018 (2016) | doi:10.1038/sdata.2016.18

Received 10 December 2015 | Accepted 12 February 2016 | Published online 15 March 2016

This article is in the 99th percentile (ranked 59th) of the 266,804 tracked articles of a similar age in all journals and the 1st percentile (ranked 1st) of the 1 tracked articles of a similar age in *Scientific Data*

EUROPEAN COMMISSION

Press Release Database

European Commission > Press releases database > Press Release details

European Commission - Statement

G20 Leaders' Communique Hangzhou Summit

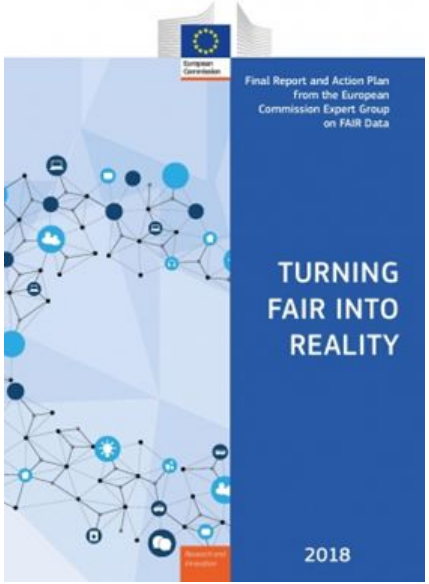
Hangzhou, 5 September 2016

1. We, the Leaders of the G20, met in Hangzhou, China on 4-5 September 2016.

G7 2017 ITALIA

Annex 4:
G7 Expert Group on Open Science

Turin, Italy, September 28, 2017



Endorsed by publishers, industry partners, funding agencies

Principles to enhance the value of all digital resources: data, software, repositories, knowledge graphs

http://www.nature.com/articles/sdata201618

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
I3. (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

http://www.nature.com/articles/sdata201618

4 facets (F,A,I,R) to the 15 principles.

Making FAIR Data

Collect

Data

Describe

Standardized Metadata

use standard metadata format

use ontologies + vocabularies

add provenance and license

Transform

Standardized Data

Use standard data format

use ontologies + vocabularies

Publish

FAIR Data

Data Repository

Persistent Metadata Identifier

Standardized Metadata

Persistent Data Identifier

Standardized Data

applies to data and their metadata

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
I3. (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

Data vs Metadata

Data are information made available for machine processing. Data can be quantitative (e.g. height, weight), qualitative (how well do you feel on a scale of 1-5), and multi-modal (text, relational, image, sound, video).

Metadata are information about data. They may contain a description, context, provenance, and meaning of the data.

Metadata

- What is the name or title of the digital resource?
- What is the digital resource about?
- Who contributed to creating or maintaining the digital resource?
- When was it created, modified, released?
- What methodology or tool was used?
- Which language is used?
- Which formats is it available as?
- What license is it released under?

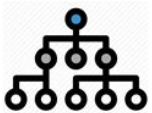
Metadata

Administrative metadata

- Provides information about the digital artifact
  - date created, size of file, type of file
  - license, permissions, etc

Descriptive metadata

- Contains information relevant to find and interpret the data
  - title, description, keywords
  - datatype-specific metadata e.g. protocol, instrument



Relational metadata

- Captures the relationship between the data item and the entity it is about or is in a contextual relation with
  - The patient for which the MRI scan was taken

Data and their metadata

Image source: https://pixabay.com/photo-782457/

Data: jpg image file

Informative metadata:

- Size: 155kb
- Date created: 2015-05-25
- Filetype: jpg

Descriptive metadata

- Title: MRI of the head
- Device: Ingenia 3.0T

Relational metadata

- About: EHR092376573
- Clinical Study: CT7812356

(meta)data for machines

The long term objective of FAIR is to make content accessible to machines, to support the everyday work we do

Data and their metadata ought to be:

- machine readable - the syntax of the data are formally specified to enable reliable reading/writing of the data.
- machine interpretable - the semantics of the data elements are well defined and can be reasoned about for information retrieval and query answering

(meta)data for machines

- find relevant digital objects that are published on the web
- facilitate **query** and **filter** content based on specific features of interest.
- easier to **understand** and **compare** data and their provenance
- easier to **replicate** experiments and **reproduce** research results
- easier to **integrate** independently produced data
- **exchange** content between different tools and environments
- **explore** and **visualize** knowledge connections

Technical infrastructure (generic operations)

Social agreements/contracts (domain-specific content)

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
I3. (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

F1: (meta) data are assigned globally unique and persistent identifiers

A unique identifier unambiguously refers a single resource, cannot be confused with another resource, and is never re-assigned to another entity.

An identifier used to denote multiple resources will confound efforts to make describes any one resource, and will lead to confusion when trying to retrieve a corresponding representation.

A persistent identifier is one where there is a technological capability and a corresponding social commitment to ensure continued access to the resource.

for instance, the responsibility to provide the resource can be transferred to another party without having to change the identifier, or there is redundancy in network to provide copies from other providers.

Globally unique and persistent?

26978244

pubmed:26978244

https://pubmed.ncbi.nlm.nih.gov/26978244/

doi:10.1038/sdata.2016.18

https://doi.org/10.1038/sdata.2016.18

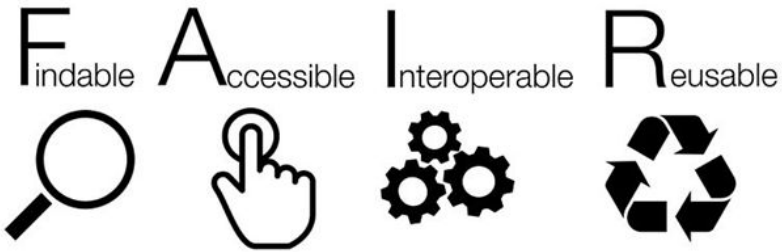
Globally unique and persistent identifiers

Obtain globally unique and persistent identifiers from a software or service that uses algorithms that can detect changes in the content.

- Persistent URLs: http://www.purlz.org
- Identifiers.org: http://identifiers.org
- Digital Object Identifier: http://www.doi.org
- Archival Resource Key (ARK): http://n2t.net/e/ark_ids.html
- Global research identifiers: https://www.grid.ac

Globally unique and persistent identifiers owing to content (via digital fingerprint)

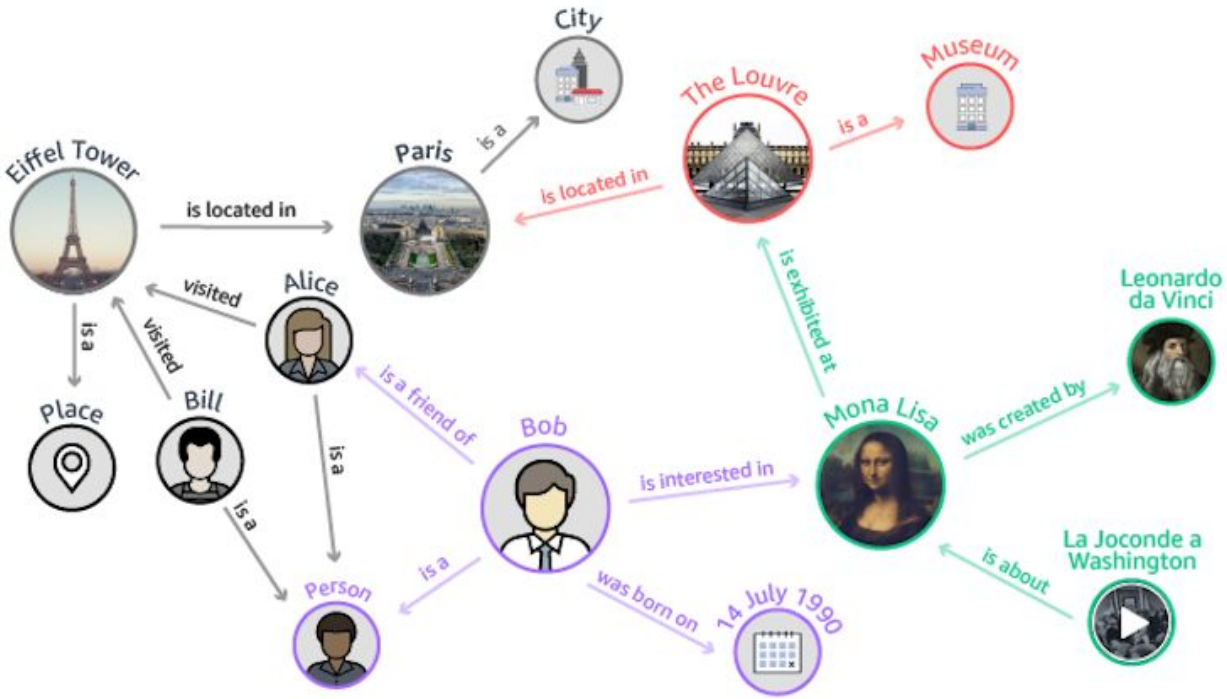
- Data GUIDs https://dataguids.org
- Trusty URIs: http://trustyuri.net/



Principles to enhance the value of *all* digital resources

data, software, web services, repositories,… ***knowledge graphs!***

Developed and endorsed by researchers, publishers, funding agencies, industry partners.


City
Museum
The Louvre
is a
Paris
is a
Eiffel Tower
is located in
is located in
is exhibited at
Leonardo da Vinci
visited
Alice
is a
visited
Place
Bill
is a friend of
Bob
Mona Lisa
was created by
is a
is interested in
is a
is a
Person
is a
was born on
14 July 1990
is about
La Joconde a Washington

slido

Which metadata could be included to describe a KG?

ⓘ Start presenting to display the poll results on this slide.

Metadata

- What is the name or title of the digital resource?
- What is the digital resource about?
- Who contributed to creating or maintaining the digital resource?
- When was it created, modified, released?
- What methodology or tool was used?
- Which language is used?
- Which formats is it available as?
- What license is it released under?
- Which descriptive or quality metrics are available?
- Who is using it?

W3C

Dataset Descriptions: HCLS Community Profile

Editors working draft.
Editors:
Alasdair J.G. Gray, Heriott-Watt University, UK <A.J.G.Gray@hw.ac.uk>
Joachim Baran, Stanford University, USA <joachim.baran@stanford.edu>
M. Scott Marshall, MAASTRO Clinic, The Netherlands <m.scott.marshall@maastro.nl>
Michel Dumontier, Stanford University, USA <michel.dumontier@stanford.edu>
Contributors:
Vladimir Alexiev, Ontotext Corp, Bulgaria <vladimir.alexiev@ontotext.com>
Peter Ansell, CSIRO, Australia <peter.ansell@csiro.au>
Gary D. Bader, The Donnelly Centre, University of Toronto, Canada <gary.bader@utoronto.ca>
Asuka Bando, NBDC, Japan <bando@biosciencedbc.jp>
Jerven Bolleman, SIB Swiss Institute of Bioinformatics, Switzerland <jerven.bolleman@isb-sib.ch>
Alison Callahan, Carleton University, Canada <alison.callahan@carleton.ca>
José Cruz-Toledo, Carleton University, Canada <josecruztoledo@cmail.carleton.ca>
Pascale Gaudet, SIB Swiss Institute of Bioinformatics, Switzerland <pascale.gaudet@isb-sib.ch>
Erich Gombocz, IO Informatics, USA <egombocz@io-informatics.com>
Alejandra Gonzalez-Beltran, University of Oxford, UK <alejandra.gonzalez.beltran@gmail.com>
Paul Groth, VU University Amsterdam, The Netherlands <p.t.groth@vu.nl>
Melissa Haendel, Oregon Health and Science University, USA <haendel@ohsu.edu>
Maori Ito, NIBIO, Japan <maori@nibio.go.jp>
Simon Jupp, EMBL-EBI, UK <jupp@ebi.ac.uk>
Nick Juty, EMBL-EBI, UK <juty@ebi.ac.uk>
Toshiaki Katayama, Database Center for Life Sciences, Japan <ktym@dbcls.jp>
Norio Kobyashi, RIKEN, Japan <norio.kobayashi@riken.jp>
Kalpana Krishnaswami, Metaome, USA <kalpana@metaome.com>
Camille Laibe, EMBL-EBI, UK <laibe@ebi.ac.uk>
Nicolas Le Novère, Babraham Institute, UK <n.lenovere@gmail.com>
Simon Lin, Marshfield Clinic Research Foundation, USA <lin.simon@mcrf.mfldclin.edu>
James Malone, EMBL-EBI, UK <malone@ebi.ac.uk>
Michael Miller, Institute for Systems Biology, USA <mmiller@systemsbiology.org>
Chris Mungall, Lawrence Berkeley National Laboratory, USA <cjm@berkeleybop.org>
Laurens Rietveld, VU University Amsterdam, The Netherlands <laurens.rietveld@vu.nl>
Sarala M. Wimalaratne, EMBL-EBI, UK <sarala@ebi.ac.uk>
Atsuko Yamaguchi, Database Center for Life Sciences, Japan <atsuko@dbcls.jp>

http://www.w3.org/TR/hcls-dataset/

A guide to describing data with RDF vocabularies

- Identifiers
- Descriptors
- Versioning
- Attribution
- Provenance
- Content summarization

Mandatory, recommended, optional descriptors
Reference editor and validation

Metadata element, description, and example of use

6.2.2 Title

At least one human-readable title should be provided for a dataset using dct:title. Alternative or older titles may be specified using dct:alternative.

For example, to provide a title and alternative title for the ChEMBL dataset:

```
:chembl
        dct:title "ChEMBL"@en ;
        dct:alternative "ChEMBLdb"@en ;
```

Basic Description

- Identifiers
- Title
- Description
- Homepage
- License
- Language
- Keywords
- Concepts and vocabularies used
- Standards
- Reference Documentation/Publication

- Format
- Download URL
- Landing page
- SPARQL endpoint

Prefix	URI	Description
cito:	http://purl.org/spar/cito/	Citation Typing Ontology
dcat:	http://www.w3.org/ns/dcat#	Data Catalog
dctypes:	http://purl.org/dc/dcmitype/	Dublin Core Metadata Types
dct:	http://purl.org/dc/terms/	Dublin Core Metadata Terms
foaf:	http://xmlns.com/foaf/0.1/	Friend-of-a-Friend
freq:	http://purl.org/cld/freq/	Collection Description Frequency Vocabulary
idot:	http://identifiers.org/idot/	Identifiers.org vocabulary
lexvo:	http://lexvo.org/ontology#	Lexical Vocabulary
pav:	http://purl.org/pav/	Provenance Authoring and Versioning ontology
prov:	http://www.w3.org/ns/prov#	PROV Ontology
rdf:	http://www.w3.org/1999/02/22-rdf-syntax-ns#	RDF Syntax
rdfs:	http://www.w3.org/2000/01/rdf-schema#	RDF Schema
schemaorg:	http://schema.org/	schema.org vocabulary
sd:	http://www.w3.org/ns/sparql-service-description#	SPARQL 1.1 Service Description
sio:	http://semanticscience.org/resource/	Semanticscience Integrated Ontology (SIO)
xsd:	http://www.w3.org/2001/XMLSchema#	XML Schema
void:	http://www.w3.org/TR/void/	Describing Linked Datasets with the VoID Vocabulary
void-ext:	http://ldf.fi/void-ext	Extensions to the Vocabulary of Interlinked Datasets (VoID)

Statistics

Statistics	
# of triples	void:triples
# of typed entities	void:entities
# of subjects	void:distinctSubjects
# of properties	void:properties
# of objects	void:distinctObjects
# of classes	void:classPartition
# of literals	void:classPartition
# of RDF graphs	void:classPartition
class frequency	void:classPartition
property frequency	void:propertyPartition
property and subject types	void:propertyPartition
property and object types	void:propertyPartition
property and literals	void:propertyPartition
property subject and object types	void:propertyPartition

*6.6.1.1 To specify the **number of triples** in the dataset:*

Pattern:

```
:rdfdataset
        void:triples "###"^^xsd:integer .
```

Example:

```
:chembl17rdf
  void:triples "409942525"^^xsd:integer .
```

SPARQL query:

```
SELECT (COUNT(*) AS ?triples)
{ ?s ?p ?o  }
```

Metagraph

Pattern:

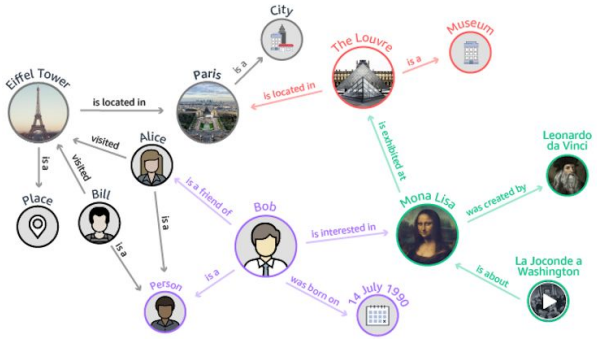
```
:rdfdataset
        void:propertyPartition [
                void:property <property-uri> ;
                void:classPartition [
                        void:class <subject-class-uri> ;
                        void:distinctSubjects "###"^^xsd:integer ;
                ];
                void-ext:objectClassPartition [
                        void:class <object-class-uri> ;
                        void:distinctObjects "###"^^xsd:integer ;
                ];
        ] .
```

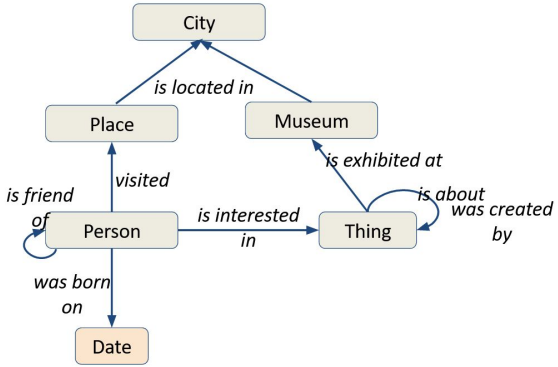
Example:

```
:chembl17rdf
    void:propertyPartition [
        void:property <http://rdf.ebi.ac.uk/terms/chembl#hasAssay> ;
        void:triples "12419715"^^xsd:integer ;
        void:classPartition [
            void:class <http://rdf.ebi.ac.uk/terms/chembl#Activity> ;
            void:distinctSubjects "12419715"^^xsd:integer ;
        ];
        void-ext:objectClassPartition [
          void:class <http://rdf.ebi.ac.uk/terms/chembl#Assay> ;
          void:distinctObjects "1042288"^^xsd:integer ;
    ]] .
```

SPARQL query:

```
SELECT (COUNT(DISTINCT ?s) AS ?scount) ?stype ?p ?otype  (COUNT(DISTINCT ?o) AS ?ocount)
{
 ?s ?p ?o .
 ?s a ?stype .
 ?o a ?otype .
} GROUP BY ?p ?stype ?otype
```





Describing Linked Datasets with the VoID Vocabulary

W3C Interest Group Note 03 March 2011

This version:
http://www.w3.org/TR/2011/NOTE-void-20110303/

Latest version:
http://www.w3.org/TR/void/

```
:DBpedia a void:Dataset;
    dcterms:title "DBPedia";
    dcterms:description "RDF data extracted from Wikipedia";
    dcterms:contributor :FU_Berlin;
    dcterms:contributor :University_Leipzig;
    dcterms:contributor :OpenLink_Software;
    dcterms:contributor :DBpedia_community;
    dcterms:source <http://dbpedia.org/resource/Wikipedia>;
    dcterms:modified "2008-11-17"^^xsd:date;
    .
:FU_Berlin a foaf:Organization;
    rdfs:label "Freie Universität Berlin";
    foaf:homepage <http://www.fu-berlin.de/>;
    .
 # Similar descriptions of the other contributors go here
```

Table of Contents

W3C Recommendation

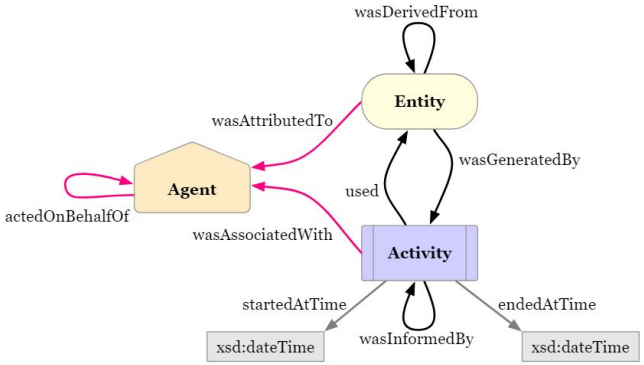
W3C

PROV-O: The PROV Ontology

W3C Recommendation 30 April 2013

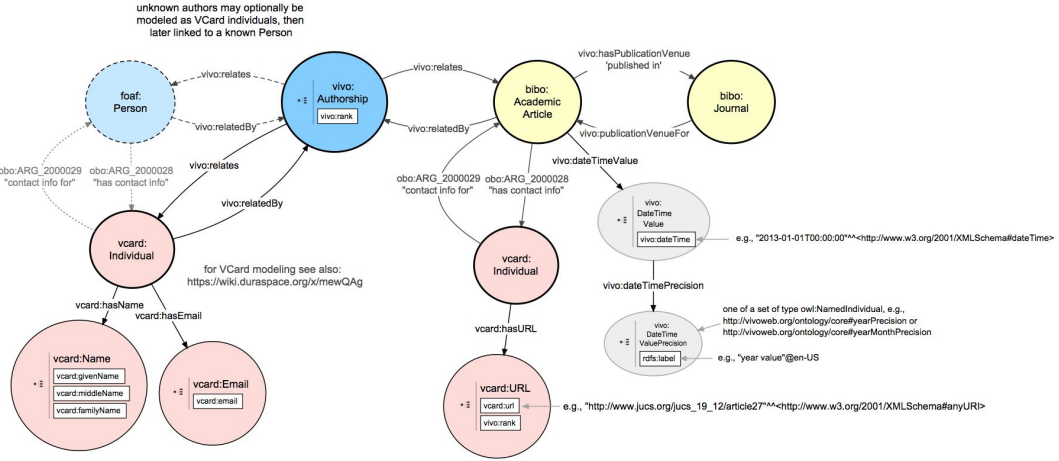
This version:
http://www.w3.org/TR/2013/REC-prov-o-20130430/
Latest published version:
http://www.w3.org/TR/prov-o/

VIVO connect share discover



8/26/15

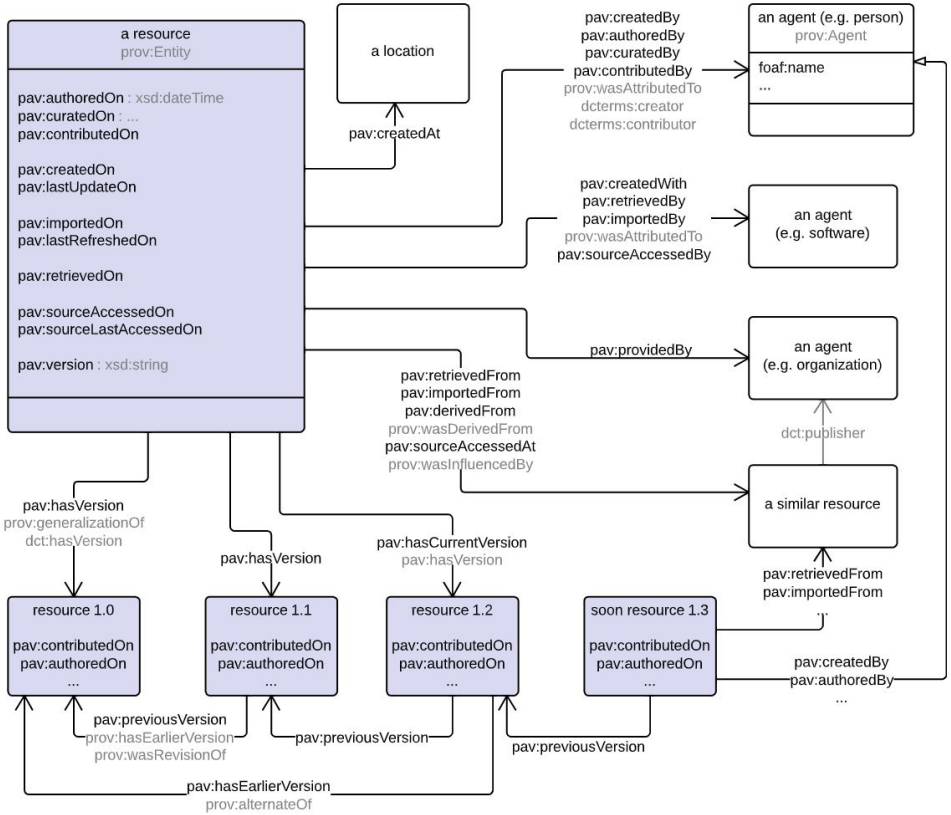
VIVO Authorship: Connecting an Author with a Publication



PAV - Provenance, Authoring and Versioning

IRI:
http://purl.org/pav/

Version IRI:
http://purl.org/pav/2.3



Bioschemas

http://bioschemas.org

Key to specification table

Schema.org properties where the Expected Types have been changed, or new (i.e., Bioschemas created) properties/types are ***green***.
Schema.org properties/types are ***red***.
Pending Schema.org properties/types are ***blue***.
External (i.e., from 3rd party ontology) properties/types are ***black***.

CD = Cardinality

Property	Expected Type	Description	CD	Controlled Vocabulary	Example
Marginality: Minimum.					
description	Text	**Schema:** A description of the item. **Bioschemas:** A short summary describing a dataset.	ONE		
identifier	PropertyValue Text URL	**Schema:** The identifier property represents any kind of identifier for any kind of Thing, such as ISBNs, GTIN codes, UUIDs etc. Schema.org provides dedicated properties for representing many of these, either as textual strings or as URL (URI) links. See background notes for more details.	MANY		
keywords	Text	**Schema:** Keywords or tags used to describe this content. Multiple entries in a keywords list are typically delimited by commas. **Bioschemas:** These keywords provide a summary of the dataset.	ONE		
name	Text	**Schema:** The name of the item. **Bioschemas:** A descriptive name of the dataset.	ONE		
rdf:type	URL	**Bioschemas:** This is used by validation tools to indentify the profile used. You must use the value specified in the Controlled Vocabulary column.	ONE		
url	URL	**Schema:** URL of the item. **Bioschemas:** The location of a page describing the dataset.	ONE		
Marginality: Recommended.					
citation	CreativeWork Text	**Schema:** A citation or reference to another creative work, such as another publication, web page, scholarly article, etc.	MANY		

Dataset

A Schema.org Type

Thing > CreativeWork > Dataset

[more...]

A body of structured information describing some topic(s) of interest.

Property	Expected Type	Description
Properties from Dataset		
distribution	DataDownload	A downloadable form of this dataset, at a specific location, in a specific format. This property can be repeated if different variations are available. There is no expectation that different downloadable distributions must contain exactly equivalent information (see also DCAT on this point). Different distributions might include or exclude different subsets of the entire dataset, for example.
includedInDataCatalog	DataCatalog	A data catalog which contains this dataset. Supersedes includedDataCatalog, catalog. Inverse property: dataset
issn	Text	The International Standard Serial Number (ISSN) that identifies this serial publication. You can repeat this property to identify different formats of, or the linking ISSN (ISSN-L) for, this serial publication.
measurementMethod	DefinedTerm or MeasurementMethodEnum or Text or URL	A subproperty of measurementTechnique that can be used for specifying specific methods, in particular via MeasurementMethodEnum.
measurementTechnique	DefinedTerm or MeasurementMethodEnum or Text or URL	A technique, method or technology used in an Observation, StatisticalVariable or Dataset (or DataDownload, DataCatalog), corresponding to the method used for measuring the corresponding variable(s) (for datasets, described using variableMeasured; for Observation, a StatisticalVariable). Often but not necessarily each variableMeasured will have an explicit representation as (or mapping to) an property such as those defined in Schema.org, or other RDF vocabularies and "knowledge graphs". In that case the subproperty of variableMeasured called measuredProperty is applicable. The measurementTechnique property helps when extra clarification is needed about how a measuredProperty was measured. This is oriented towards scientific and scholarly dataset publication but may have broader applicability; it is not intended as a full representation of measurement, but can often serve as a high level summary for dataset discovery. For example, if variableMeasured is: molecule concentration, measurementTechnique could be: "mass spectrometry" or "nmr spectroscopy" or "colorimetry" or "immunofluorescence". If the variableMeasured is "depression rating", the measurementTechnique could be "Zung Scale" or "HAM-D" or "Beck Depression Inventory". If there are several variableMeasured properties recorded for some given data object, use a PropertyValue for each variableMeasured and attach the corresponding measurementTechnique. The value can also be from an enumeration, organized as a MeasurementMetholdEnumeration.
variableMeasured	Property or PropertyValue or StatisticalVariable or Text	The variableMeasured property can indicate (repeated as necessary) the variables that are measured in some dataset, either described as text or as pairs of identifier and description using PropertyValue, or more explicitly as a StatisticalVariable.
Properties from CreativeWork		
about	Thing	The subject matter of the content. Inverse property: subjectOf
abstract	Text	An abstract is a short description that summarizes a CreativeWork.

Google Dataset Search

Search for Data Sets

WIKIMEDIA

Wikidata

Explore at www.wikidata.org | Explore at m.wikidata.nym.sk | Explore at unblocked.to

Explore at the Datahub | Explore at data.wu.ac.at

6 scholarly articles cite this data set (View in Google Scholar)

Data set provided by

Wikimedia Foundation

Licence

CC0 1.0 Universal Public Domain Dedication

Licence information was derived automatically

Available download formats from providers

sparql endpoint, full rdf turtle dump, full json dump, simplified ("truthy") rdf n-triples dump

Description

Wikidata offers a wide range of general data about our universe as well as links to other databases. The data is published unde the CC0 "Public domain dedication" license. It can be edited by anyone and is maintained by Wikidata's editor community.


```
{
  "@context": "http://schema.org",
  "@id": "http://datahub.io/dataset/e2427d01-3ef6-40f7-a148-f0bf28176184",
  "@type": "Dataset",
  "dateModified": "2015-11-11T10:59:55.856901",
  "datePublished": "2015-11-11T10:18:28.306264",
  "description": "The free knowledge base anyone can edit\r\nhttps://wikidata.org",
  "distribution": [
    {
      "@id": "http://datahub.io/dataset/e2427d01-3ef6-40f7-a148-f0bf28176184/resource/8fabf551-3d9e-4e23-93cb-0ef03ce45fea",
      "@type": "DataDownload",
      "contentUrl": "https://www.wikidata.org/w/api.php",
      "datePublished": "2015-11-11T10:21:42.576076",
      "description": "The MediaWiki action API is a web service that provides convenient access to wiki features, data, and meta-data over HTTP, via a URL at api.php. ",
      "fileFormat": "api/json",
      "license": "http://www.opendefinition.org/licenses/cc-zero",
      "name": "Wikidata API"
    },
```

Google Structured Data Testing Tool

https://data.wu.ac.at/schema/datahub_io/ZTI0MjdkMDEtM2VmNi00MGY3LWExNDgtZjBiZjI4MTc2MTg0

NEW TEST

```
            .ui.card .meta {
                font-size: 0.8em;
            }
        </style>
    </head>
    <body>
        <header id="header" class="ui vertical segment">
            <h2 class="ui header">
                <div class="content">
                    Open Data Portal Watch
                    <div class="sub header">Mapping and export of Schema.org metadata descriptions for over 250 Open Data p
                </div>
                <a href="http://data.wu.ac.at/"><img  class="ui small image" src="http://data.wu.ac.at/logo_b.png"></a>
            </h2>
        </header>
        <nav id="mainnav" class="ui fluid secondary stackable pointing icon menu background">
            <div class="ui item white"><a class="ui button" href="http://data.wu.ac.at/schema/">Data Portal List</a></div>
            <div class="ui item white"><a class="ui primary button" href="http://data.wu.ac.at/odgraphsearch/">Spatio-Tempo
            <div class="ui item white"><a class="ui primary button" href="http://data.wu.ac.at/portalwatch/">Data Portal Mo
            <!--<a class="link item"><i class="icon download"></i>Data</a>-->
            <div class="right menu">
                <div class="ui item white"><a href="http://data.wu.ac.at/portalwatch/about" class="ui primary button">About
            </div>
        </nav>
        <!-- CONTENT -->
        <section class="ui text">
            <div class="ui container">
                    <div class="ui two column grid">
                        <div class="row">
                            <div class="column">
                                <div class="ui fluid card">
                                    <div class="content">
                                        <div class="header">
                                            Dataset Information
                                        </div>
                                        <div class="description">
                                            <table class="ui very basic red compact table">
                                                <tbody>
                                                    <tr>
                                                        <td class="right aligned">Title</td>
                                                        <td><a href="http://datahub.io/dataset/e2427d01-3ef6-40f7-a148-f0bf
                                                    </tr>
                                                    <tr>
                                                        <td class="right aligned">Description</td>
                                                        <td>The free knowledge base anyone can edit
```

Dataset

All (1)

Dataset

0 ERRORS 3 WARNINGS

ID: http://datahub.io/dataset/e2427d01-3ef6-40f7-a148-f0bf28176184

@type	Dataset
@id	http://datahub.io/dataset/e2427d01-3ef6-40f7-a148-f0bf28176184
dateModified	2015-11-11T10:59:55
datePublished	2015-11-11T10:18:28
description	The free knowledge base anyone can edit https://wikidata.org
keywords	Wikimedia
keywords	wikibase
keywords	Wikidata
keywords	open data
name	Wikidata
distribution	
@type	DataDownload
@id	http://datahub.io/dataset/e2427d01-3ef6-40f7-a148-f0bf28176184/resource/8fabf551-3d9e-4e23-93cb-0ef03ce45fea
contentUrl	https://www.wikidata.org/w/api.php
datePublished	2015-11-11T10:21:42
description	The MediaWiki action API is a web service that provides convenient access to wiki features, data, and meta-data over HTTP, via a URL at api.php.
fileFormat	api/json
license	http://www.opendefinition.org/licenses/cc-zero
name	Wikidata API
⚠ **encodingFormat**	The *encodingFormat* field is recommended. Please provide a value if available.
distribution	

FAIRsharing.org
standards, databases, policies

databases > doi:10.25504/fairsharing.6s749p

Dev Wikidata

General Information

Free knowledge database project hosted by Wikimedia and edited by volunteers.

Homepage http://wikidata.org/
Developed in Worldwide
Created in 2012

FAIRsharing is a repository of standards and databases. It exposes resource metadata using schema.org


```
<script type="application/ld+json">
{
        "@context": "http://schema.org",
        "@id": "https://doi.org/10.25504/FAIRsharing.6s749p",
        "@type": "Dataset",
        "alternateName": "Wikidata",
        "citation": {
                "@type": "CreativeWork",
                "identifier": "doi:10.25504/FAIRsharing.6s749p"
        },
        "datePublished": "2016-09-06 16:10:14",
        "description": "Free knowledge database project hosted by Wikimedia and edited by volunteers.",
        "identifier": "10.25504/FAIRsharing.6s749p",
        "license": "https://creativecommons.org/licenses/by-sa/4.0/. Please link to https://fairsharing.org and
https://fairsharing.org/static/img/home/svg/FAIRsharing-logo.svg for attribution.",
        "name": "Wikidata",
        "url": "https://doi.org/10.25504/FAIRsharing.6s749p"
}
```

Publishing RDF - Linked Data Principles

1. Use **Uniform Resource Identifiers (URIs/URLs)** as identifiers for things
2. Use **HTTP URIs**, so that people can look up those entities
3. When someone looks up a URI, provide **useful information**, using Semantic Web standards
4. Include **links** to other URIs, so that they can discover more things

FAIR Knowledge Graphs

1. **Build your knowledge graph using existing standards**
    a. Assign a **unique identifier (URI)** to every entity (kg, types, relations, instances)
    b. Format the data using a **data standard** (e.g. RDF, nanopublications)
    c. Capture the **provenance** and **context** of each assertion, and of the graph itself
    d. Where possible, make **links** to other published resources
2. **Create high quality metadata to document your KG**
    a. **use** metadata schemas (schema.org, HCLS)
    b. **add KG specific metadata** (eg. number of triples, vocabs used, etc)
    c. **validate** the metadata against the schema
3. **Make your KG available to others**
    a. **Version** your KG, and **publish the data and metadata** to an archive
    b. Create programmatic access points via APIs

Publish to FAIR Repositories

Hugging Face publishes structured metadata and has APIs to access the content. Also can support large files through git LFS.

Hugging Face | Search models, datasets, users... | Models | Datasets

Datasets: KGraph / FB15k-237 | like 4

Tasks: Other | Languages: English | Multilinguality: monolingual | Size Categories: 100K<n<1M | Annotations Creators: found | crowdsourced

Tags: knowledge graph | knowledge | link prediction | link | Croissant | License: cc-by-4.0

Dataset card | Viewer | Files and versions | Community

Dataset Viewer | Auto-converted to Parquet | API | View in Dataset Viewer

Split (3)
train · 272k rows

▸ The full dataset viewer is not available (click to read why). Only showing a preview of the rows.

text string
/m/0rh6k /location/statistical_region/religions./location/religion_percentage/religion /m/01lp8
/m/05d7rk /award/award_nominee/award_nominations./award/award_nomination/award /m/027dtxw
/m/05hs4r /music/genre/artists /m/01pbxb
/m/03qcq /influence/influence_node/influenced_by /m/084w8
/m/01lxd4 /music/genre/artists /m/0f0y8
/m/08815 /education/educational_institution/students_graduates./education/education/student /m/05bnp0
/m/0g56t9t /film/film/release_date_s./film/film_regional_release_date/film_release_region /m/05r4w
/m/0160w /organization/organization_member/member_of./organization/organization_membership/organization /m/02vk52z
/m/0dbpyd /people/person/place_of_birth /m/09c7w0
/m/053y0s /people/person/profession /m/0dz3r

‹ Previous | 1 | 2 | 3 | ... | 2,722 | Next ›

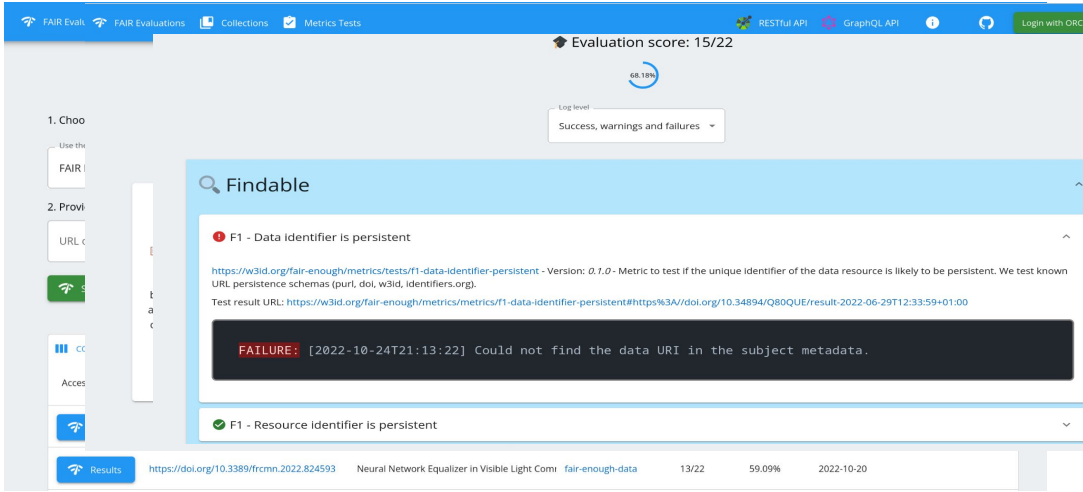
Dataset Card for FB15k-237

Dataset Summary

FB15k-237 is a link prediction dataset created from FB15k. While FB15k consists of 1,345 relations, 14,951 entities, and 592,213 triples, many triples are inverses that cause leakage from the training to testing and validation splits. FB15k-237 was created by Toutanova and Chen (2015) to ensure that the testing and evaluation datasets do not have inverse relation test leakage. In summary, FB15k-237 dataset contains 310,079 triples with 14,505 entities and 237 relation types.

Quality of FAIR Implementation

- **FAIR Enough** is a system to perform automated assessment of the technical quality of the FAIRness implementation.
- FAST and can be used with different metric collections
- Keeps track of past assessments to monitor status
- Extensible via service based framework (can use FAIR Evaluator harvester and metrics)
- Open source and Docker deployable



https://fair-enough.semanticscience.org

Hugging Face | Search models, datasets, users... | Models | Datasets | Spaces | Docs | Solutions | Pricing

meta-llama / **Llama-2-70b-hf** | like 369

Text Generation | PyTorch | Safetensors | Transformers | English | llama | facebook | meta | llama-2 | text-generation-inference | arxiv:2307.09288

Model card | Files and versions | Train | Deploy | Use in Transformers

Access Llama 2 on Hugging Face

This is a form to enable access to Llama 2 on Hugging Face after you have been granted access from Meta. Please visit the Meta website and accept our license terms and acceptable use policy before submitting this form. Requests will be processed in 1-2 days.

Downloads last month
168,220

Safetensors | Model size | 69B params | Tensor type | F16 · F32

https://huggingface.co/meta-llama/Llama-2-70b-hf

Identifier of this evaluation: https://w3id.org/fair-enough/evaluations/2d2ff90ff23f8ee901295b5786a950e4ddbf0b16

Evaluated in 3s with the fair-enough-metadata collection on the 2023-07-26

By https://orcid.org/0000-0003-4727-9435

See other evaluations for https://huggingface.co/meta-llama/Llama-2-70b-hf

Extracted metadata

Title: meta-llama/Llama-2-70b-hf · Hugging Face

Evaluation score: 11/16