

Publicación de datos FAIR

Mikel Egaña Aranguren

mikel-egana-aranguren.github.io

mikel.egana@ehu.eus



Publicación de datos FAIR

<https://github.com/mikel-egana-aranguren/UM-Bioinformatics-MSc-FAIR-data>



Publicación de datos FAIR

```
git clone https://github.com/mikel-egana-aranguren/UM-Bioinformatics-MSc-FAIR-data.git
```

Índice

1. Introducción
2. Principios FAIR
3. Publicar datos FAIR
4. Recursos sobre FAIR
5. Linked Data
6. Ejemplo práctico
7. Proyecto a realizar

Introducción

Principios FAIR: una mejor publicación de datos (Científicos)

Para humanos y **máquinas**

No es un estándar

No promueven una tecnología concreta

Introducción

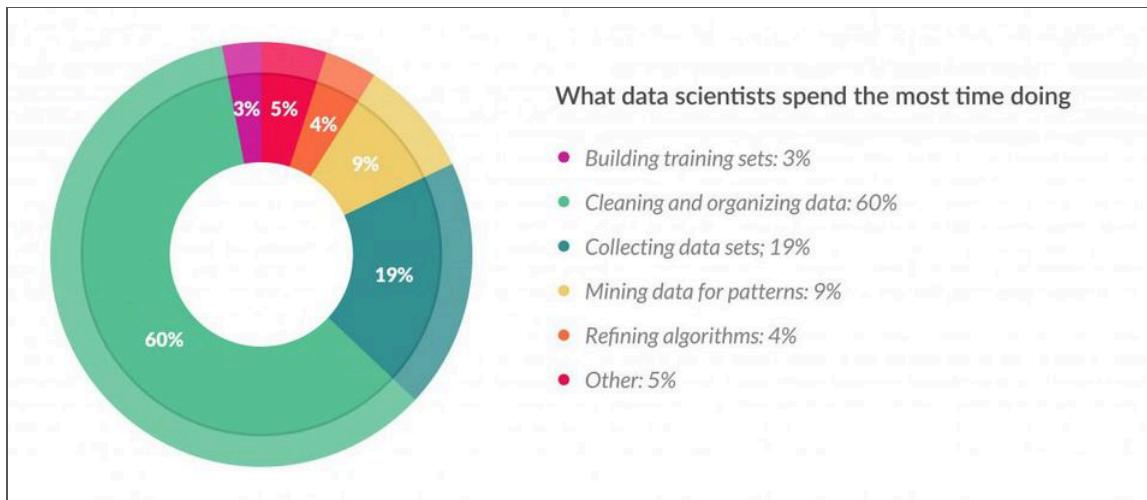
Son principios **guía**

No se cumplen de manera binaria (aprobado o no)

Un sistema siempre puede ser "más FAIR"

Introducción

80% del tiempo buscando, filtrando, masajeando e integrando datos



["Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says" \[FORBES, 2020-11-19\]](#)

Introducción

La reproducibilidad es **crucial** en ciencia:

- Reproducir: ejecutar un experimento/estudio con los mismos datos/materiales
- Replicar: ejecutar un experimento/estudio con nuevos datos/materiales

Introducción

Crisis de la reproducibilidad debido a:

- Datos no publicados
- Datos publicados de manera inadecuada

Introducción

Principios FAIR para una mejor publicación de **(meta)datos**

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016)

Introducción

Cada vez más agencias gubernamentales exigen cumplir los principios FAIR a la hora de publicar los resultados científicos para recibir financión (Data Management Plans): Horizon Europe, OpenAIRE, etc.

Open Data Europa para medir la calidad de los metadatos

Iniciativas del Espacio Europeo de Datos como GAIA-X

Introducción

Ley Orgánica del Sistema Universitario (LOSU): artículo 12 (Fomento de la Ciencia Abierta y Ciencia Ciudadana)

Estrategia Nacional de Ciencia Abierta 2023-2027 (ENCA) como objetivo estratégico

Grandes empresas como Novartis, Bayer, BASF, SIEMENS ENERGY etc. usan principios FAIR para publicación interna de datos

Principios FAIR

Findable

Accesible

Interoperable

Reusable

Findable

F1. (Meta)Data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (R1)

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

F4. (Meta)Data are registered or indexed in a searchable resource

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

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

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

Principios y ejemplos

Las tecnologías para implementar los principios FAIR no son los principios FAIR

Findable

Data should be identified using globally unique, resolvable, and persistent identifiers, and should include machine-actionable contextual information that can be indexed to support human and machine discovery of that data

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

Globally unique

Dominios como um.es (€)

Registros

Algoritmos ([UUID](#))

etc.

Persistent

Infraestructura propia (€€€): por ejemplo [W3C URI Persistence Policy](#)

Registros: [identifiers.org](#), [DOI](#), [Orcid](#), [Zenodo](#), etc.

HTTP URIs

URI: Uniform Resource Identifier ([RFC 3986](#))

Identifica un recurso (URL: Localiza un documento)

HTTP: podemos usar HTTP para acceder (**Resolver**) a esa URI (dominio)

Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data

McMurry JA, Juty N, Blomberg N, Burdett T, Conlin T, et al. (2017) PLOS Biology 15(6): e2001414. <https://doi.org/10.1371/journal.pbio.2001414>

[papers/journal.pbio.2001414.pdf](#)

F1. Ejemplos

- <https://orcid.org/0000-0001-8888-635X>
- **doi:**10.4121/uuid:5146dd06-98e4-426c-9ae5-dc8fa65c549f -
<https://doi.org/10.4121/UUID:5146DD06-98E4-426C-9AE5-DC8FA65C549F>
- <http://www.uniprot.org/uniprot/P98161>
- <http://omim.org/entry/173900>

F1. Nuestro ejemplo hipotético

Nuestro laboratorio de la UM ha descubierto un gen nuevo, PKD1, implicado en enfermedades renales de los humanos

La UM tiene un repositorio persistente de datos

URI del dataset: <https://um.es/dataset/UMGenesDataset>

URI de un gen: <https://um.es/data/LDD773322>

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

Añadir metadatos lo más detallados posible

Metadatos de contenido: a qué especies pertenecen los genes, la temática de los datos, etc.

Metadatos técnicos: cuándo se generaron los datos, como, por quién, etc.

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

Se usan ontologías (I1)

Repositorios de ontologías: [Linked Open Vocabularies](#), [OBO Foundry](#),
[BioPortal](#), [BioSchemas](#), etc.

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

"<https://um.es/dataset/UMGenesDataset> was generated on 2020-12-10T13:00:07"

"<https://um.es/dataset/UMGenesDataset> is about genes"

"<https://um.es/dataset/UMGenesDataset> relates to humans"

etc.

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

Repositorios generales: [Zenodo](#), [DataDryad](#), [Dataverse](#) ([Harvard Dataverse](#)), etc.

Repositorios temáticos: [UniProt](#), [GenBank](#), etc.

Indexadores como Google

Indexación

Google indexa de manera "básica" ...

... pero cada vez menos, gracias a [Schema](#) (Ontología muy ligera para describir datos en la web) y [JSON-LD: Bioschemas](#)

Hay que intentar publicar buenos metadatos para una indexación adecuada (por Google o cualquier agente que *entienda* las ontologías que usamos)

Accessible

Identified data should be accessible, optimally by both humans and **machines**, using a clearly-defined protocol and, if necessary, with clearly-defined rules for authorization/authentication

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

Por ejemplo [HTTPS](#)

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

Por ejemplo [HTTPS](#)

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

Hacer explícitas las condiciones físicas de acceso, para humanos y **máquinas**

Datos protegidos por propiedad intelectual o privacidad (Ej. datos clínicos): no se publican los datos, sólo (algunos) metadatos y sus condiciones de acceso

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

Conservar datos es muy caro

Conservar metadatos es mucho más barato

Si los datos ya no existen, deberíamos ser explícitos sobre ello, por ejemplo para evitar búsquedas innecesarias

Interoperable

Data becomes interoperable when it is machine-actionable, using shared vocabularies and/or ontologies, inside of a syntactically and semantically machine-accessible format

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for Knowledge Representation

Las máquinas también tienen que *entender* los (meta)datos

Por ejemplo OWL ([Web Ontology Language](#))

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

Las ontologías usadas para describir los datos también se tienen que publicar siguiendo los principios FAIR

I3. (Meta)data include qualified references to other (meta)data

Los (meta)datos son solo útiles cuando los integramos con otros datos

Enlaces explícitos a otros datos: "part-of", "catalyses", etc.

Las máquinas entienden el significado de esa relación

Reusable

Reusable data will first be compliant with the F, A, and I principles, but further, will be sufficiently well-described with, for example, contextual information, so it can be accurately linked or integrated, like-with-like, with other data sources. Moreover, there should be sufficiently rich provenance information so reused data can be properly cited

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

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

F2 es para descubrir datos, R1 es para decidir si los datos son útiles

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

Describe the scope of your data: for what purpose was it generated/collected?

Mention any particularities or limitations about the data that other users should be aware of

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

Specify the date of generation/collection of the data, the lab conditions, who prepared the data, the parameter settings, the name and version of the software used

Is it raw or processed data?

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

Ensure that all variable names are explained or self-explanatory (i.e., defined in the research field's controlled vocabulary)

Clearly specify and document the version of the archived and/or reused data

R1.1. (Meta)data are released with a clear and accessible data usage license

I es sobre interoperabilidad técnica; R1.1 es sobre interoperabilidad legal

Los datos deben tener una licencia clara y explícita para humanos y máquinas

Por ejemplo, [Creative Commons RDF](#)

R1.2. (Meta)data are associated with detailed provenance

¿Cómo, quién, cuándo, por qué generó los datos?

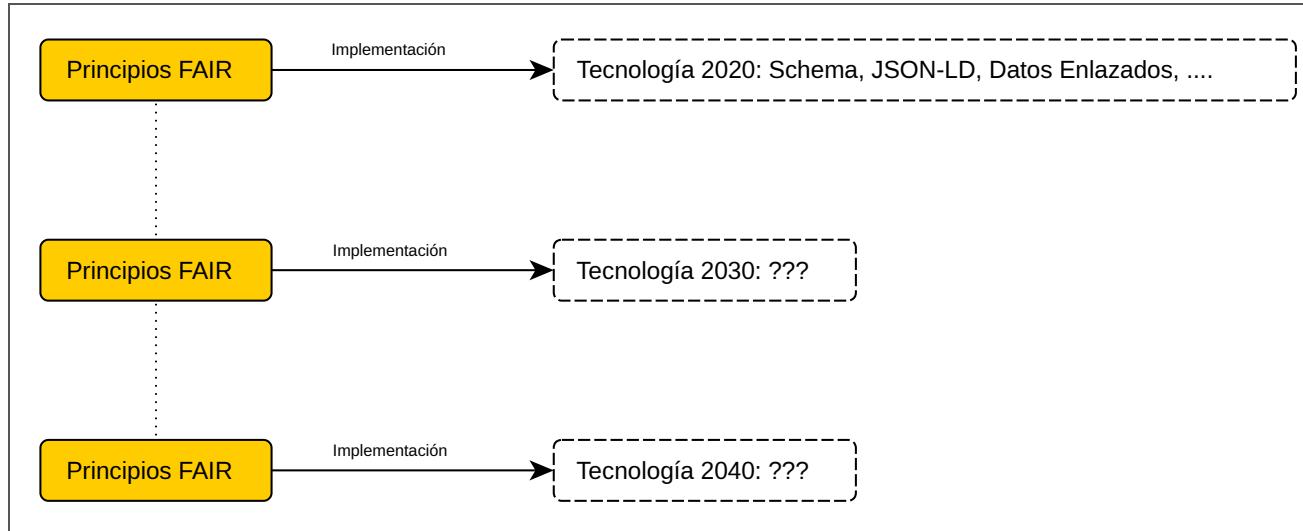
[PROV-O: The PROV Ontology](#)

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

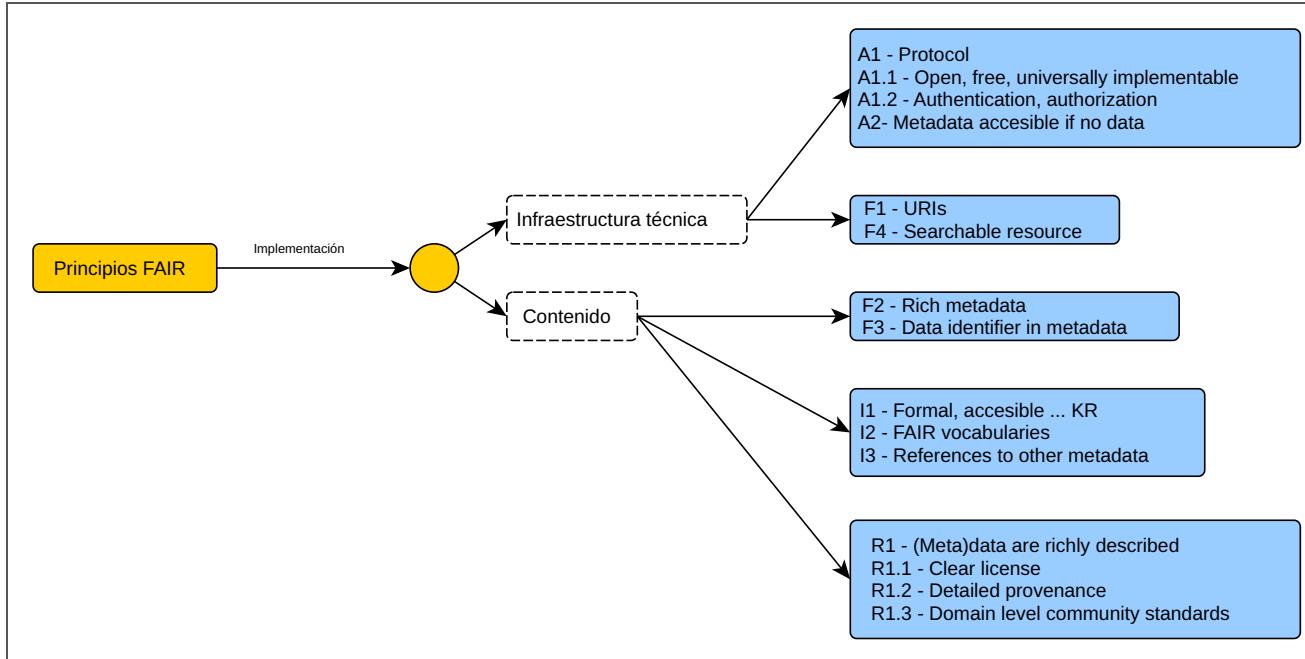
Respetar las buenas prácticas, estándares, vocabularios etc. de la comunidad científica que trabaja con esos datos

Por ejemplo [FAIR Sharing Standards](#)

Principios FAIR vs implementación



Principios FAIR: implementación



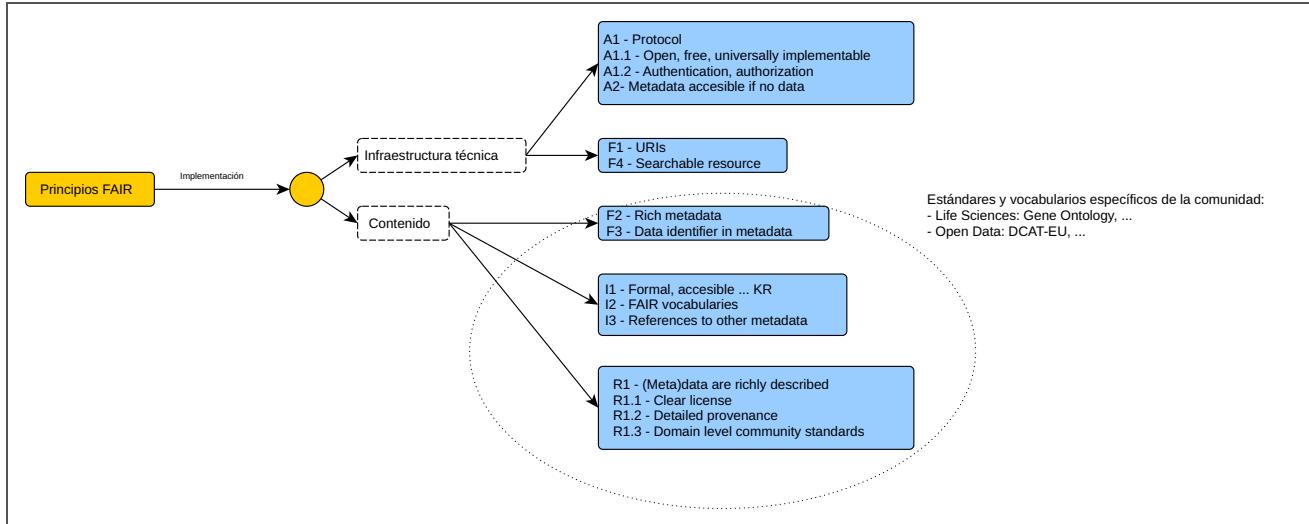
Madurez FAIR

A design framework and exemplar metrics for FAIRness

<https://github.com/FAIRMetrics/Metrics>

FAIR Evaluation Services

Madurez FAIR



Proyectos, Empresas, y centros

[Personal Health Train Network \(1812.00991.pdf\)](#)

[FAIR Data Systems](#)

[The Hyve](#)

[Eccenca GmbH](#)

[Dutch Tech Centre for Life Sciences \(DTL\)](#)

Recursos sobre FAIR en internet

[The FAIR cookbook](#)

[GO FAIR foundation](#)

[FAIR Sharing](#)

[FAIR-DOM](#)

Publicacion datos FAIR

Hay muchas maneras de publicar datos siguiendo los principios FAIR, dependiendo de la tecnología: API REST, Linked Data, Linked Data Fragments, FAIR Data Point, ...

Estas soluciones se ocupan de la parte *técnica*

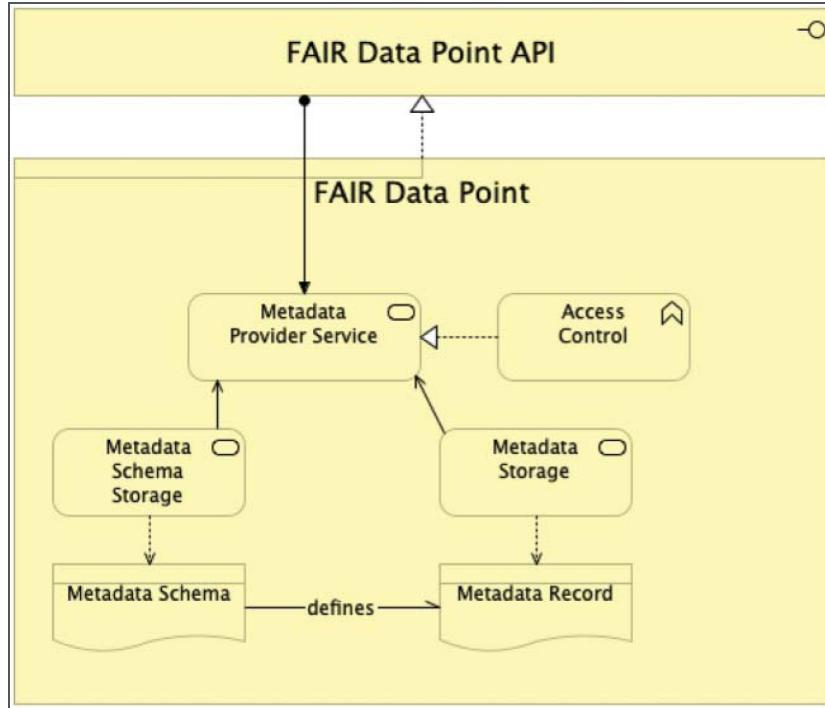
Pero no suficiente: hay que producir contenido FAIR (Metadatos, Ontologías, URIs, etc.)

FAIR Data Point

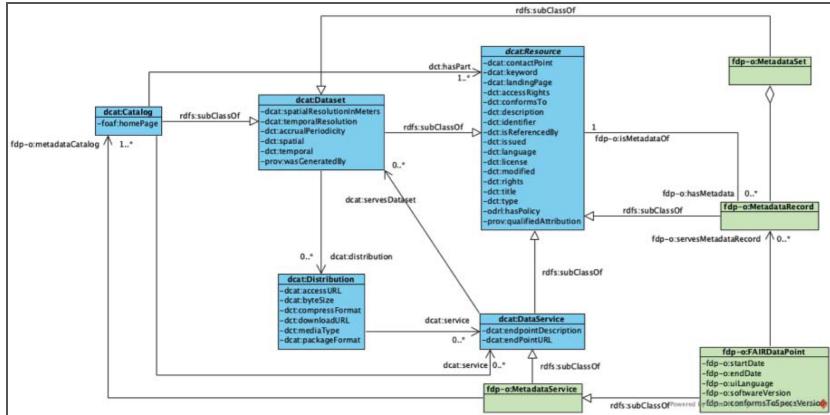
FDP

FAIR Data Point: A FAIR-Oriented Approach for Metadata Publication

FAIR Data Point



FAIR Data Point



FAIR Data Point

FAIR FAIR Data Point

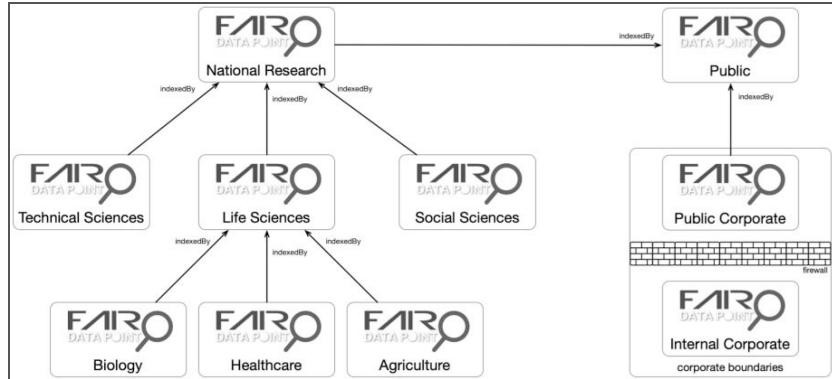
Search FAIR Data Point... Log in

FAIR Data Points

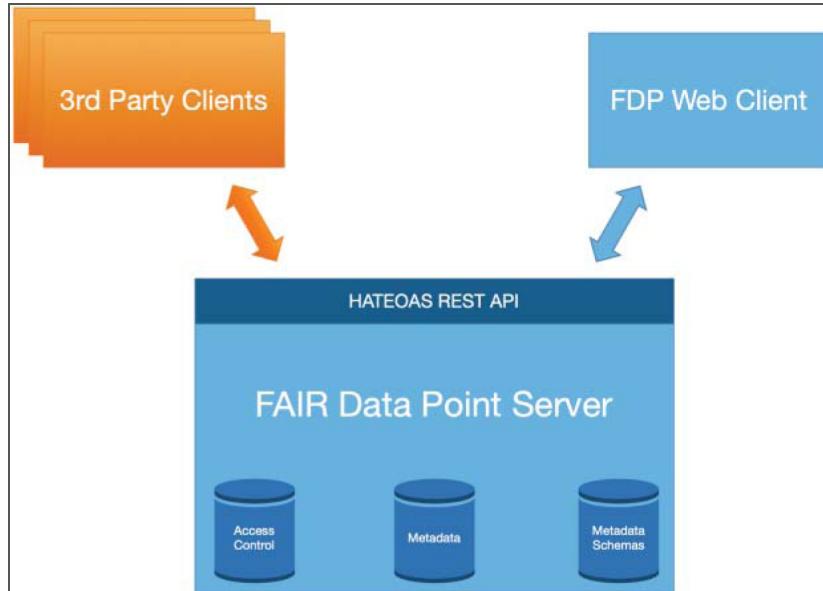
Filter: All 167 Active 24 Inactive 33 Unreachable 77 Invalid 30 Unknown 3

Endpoint	Registration	Modification	Status
https://app.fairdatapoint.org	29-04-2020, 16:37:21	21-02-2022, 16:47:21	ACTIVE
https://fdp.sdsc.edu	01-05-2020, 23:44:58	23-02-2022, 04:03:04	ACTIVE
https://fdp.umcn.nl	26-08-2020, 14:58:14	21-02-2022, 11:53:35	ACTIVE
https://home.fairdatapoint.org	20-01-2021, 21:56:42	21-02-2022, 21:25:51	ACTIVE
https://directory.bbmri-eric.eu/api/fdp	27-01-2021, 15:30:32	25-02-2022, 05:45:05	ACTIVE
https://covid19research.nl/api/fdp	27-01-2021, 15:39:43	25-02-2022, 05:00:05	ACTIVE
https://fdp.castoredc.com	01-02-2021, 11:57:53	25-02-2022, 07:00:02	ACTIVE
https://twoc.fair-dtis.surf-hosted.nl	04-03-2021, 11:32:52	22-02-2022, 08:50:19	ACTIVE
https://diamonds.tno.nl/fairdatapoint	12-05-2021, 13:35:58	19-02-2022, 08:51:04	ACTIVE
https://w3id.org/duchenne-fdp	19-05-2021, 09:54:04	22-02-2022, 07:51:42	ACTIVE
https://w3id.org/orphadata/fairdatapoint	27-05-2021, 17:09:53	24-02-2022, 10:07:08	ACTIVE
https://twoc-index.fair-dtis.surf-hosted.nl	09-06-2021, 14:24:34	24-02-2022, 10:19:28	ACTIVE
https://w3id.org/ejp-rd/fairdatapoints/bbmri	15-06-2021, 18:44:46	21-02-2022, 14:03:09	ACTIVE
https://w3id.org/ejp-rd/fairdatapoints/wp13	23-06-2021, 12:34:47	21-02-2022, 14:00:44	ACTIVE
https://fdp.cmbi.umcn.nl	27-07-2021, 11:51:19	24-02-2022, 10:33:30	ACTIVE

FAIR Data Point



FAIR Data Point



FAIR Data Point

The screenshot shows the FAIR Data Point application interface. At the top left is the FAIR Data Point logo. A search bar and a dropdown menu labeled 'LB' are at the top right. A red ribbon banner is positioned in the top right corner.

Demonstration FAIR Data Point

This FAIR Data Point deployment is used for demonstration of the application and to allow the navigation through its metadata content. The metadata presented here is also for demonstration purposes only and not necessarily describe properly their related resources.

Catalogs

+ Create

- COVID-19 dataset catalog**
A catalog containing the metadata of a number of COVID-19-related datasets.
www.vodan-totafica.info SIO_001410
Issued 05-06-2020 Modified 28-10-2020
- COVID-19 websites catalog**
A catalog listing the metadata of a number of websites providing information about differences aspects of the COVID-19 pandemic.
Issued 05-06-2020 Modified 30-09-2020
- Example UT Data Archive catalog**
Test
synthetic
Issued 16-07-2020 Modified 16-07-2020
- FAIR Data Points catalog**
A catalog listing the metadata of a number of deployments of the FAIR Data Point.
Issued 05-06-2020 Modified 05-06-2020
- FAIR semantics catalog**
A catalog listing the metadata of ontologies relevant to the FAIR principles.
Issued 05-06-2020 Modified 08-12-2021

Metadata Issi
29-05-2020

Conforms to

- fdpMetad
- Repository

Version
1.0

Language
English

License
[cc-by-nc-nd4.0](#)

Start date
01-06-2020

Institution country
Q55

Download RDF
ttl rdf+xml json-lld

LB

FAIR Data Point

- Users
- Resources definitions
- SHACL shapes
- Settings
- Reset to defaults

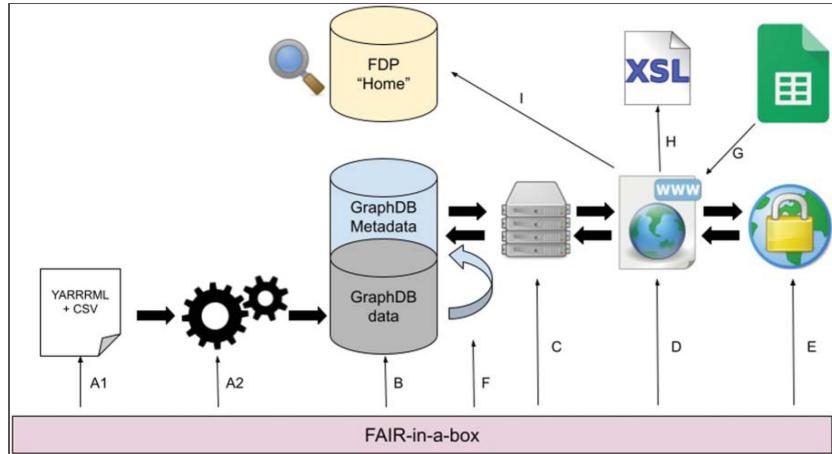
Luiz Bonino

- My Metadata
- API Keys
- Edit profile
- Log out

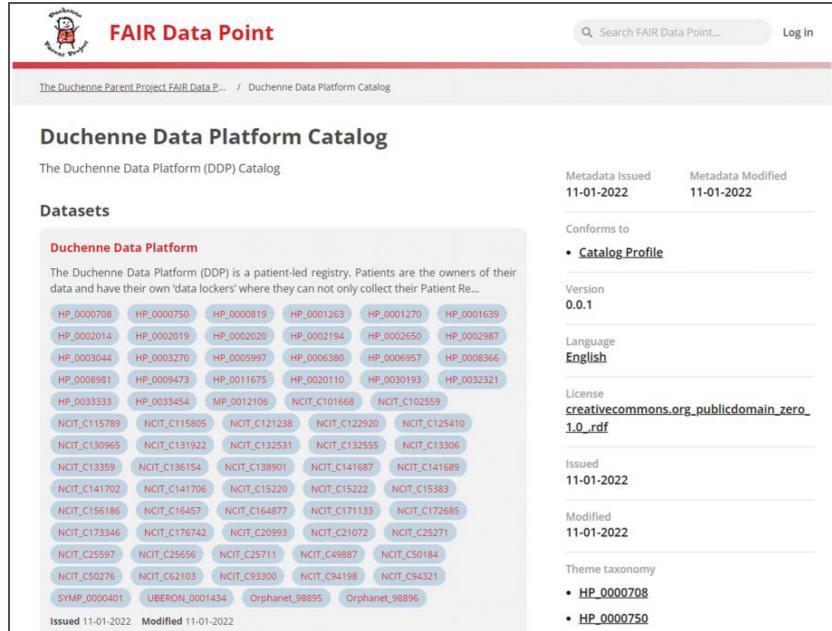
FAIR Data Point

The FAIR Data Point: Interfaces and Tooling

FAIR Data Point



FAIR Data Point

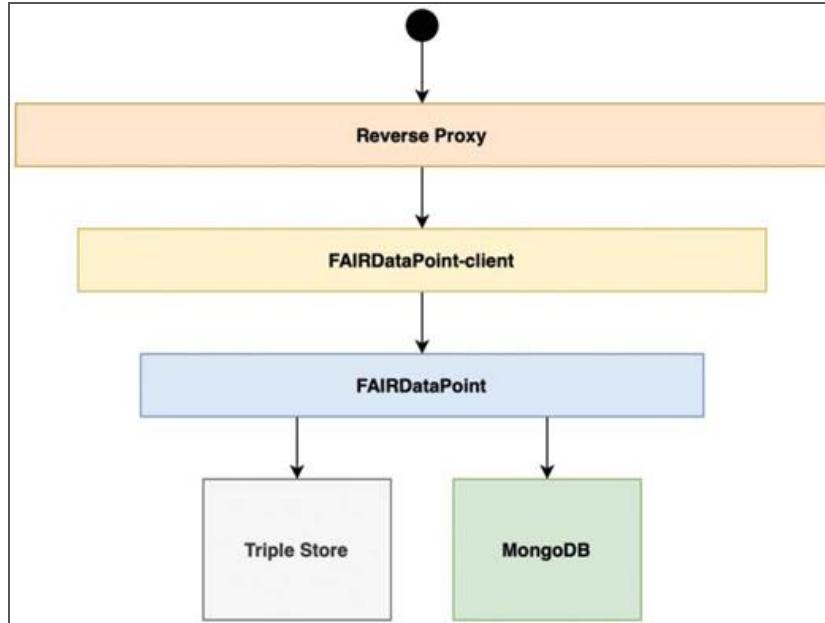


The screenshot shows the FAIR Data Point website interface. At the top, there is a logo of a person with a stethoscope, a search bar with the placeholder "Search FAIR Data Point...", and a "Log in" button. Below the header, a navigation bar includes links for "The Duchenne Parent Project FAIR Data P..." and "Duchenne Data Platform Catalog". The main content area is titled "Duchenne Data Platform Catalog" and describes it as "The Duchenne Data Platform (DDP) Catalog". A section titled "Datasets" contains a table with two columns of dataset identifiers. To the right of the table, detailed metadata is listed under sections like "Conforms to", "Version", "Language", "License", "Issued", "Modified", and "Theme taxonomy". At the bottom left, the date "Issued 11-01-2022" and "Modified 11-01-2022" is shown.

Dataset Identifier	Dataset Identifier
HP_0000708	HP_0000750
HP_0000819	HP_0001263
HP_0001270	HP_0001639
HP_0002014	HP_0002019
HP_0002020	HP_0002194
HP_0002650	HP_0002967
HP_0003044	HP_0003270
HP_0005997	HP_0006380
HP_0006957	HP_0008366
HP_0008981	HP_0009473
HP_0011675	HP_0020110
HP_0030193	HP_0032321
HP_0033333	HP_0033454
MP_0012106	NCIT_C101668
NCIT_C102559	NCIT_C102559
NCIT_C115789	NCIT_C115805
NCIT_C121288	NCIT_C12290
NCIT_C125410	NCIT_C132531
NCIT_C130965	NCIT_C131922
NCIT_C132555	NCIT_C13306
NCIT_C13359	NCIT_C136154
NCIT_C138901	NCIT_C141687
NCIT_C141689	NCIT_C141702
NCIT_C141706	NCIT_C15220
NCIT_C15222	NCIT_C15383
NCIT_C156186	NCIT_C16457
NCIT_C164877	NCIT_C171133
NCIT_C172685	NCIT_C21072
NCIT_C173346	NCIT_C176742
NCIT_C20993	NCIT_C25271
NCIT_C25597	NCIT_C25656
NCIT_C25711	NCIT_C49887
NCIT_C50184	NCIT_C50276
NCIT_C62103	NCIT_C93300
NCIT_C94198	NCIT_C94321
SYMP_0000401	UBERON_0001434
Orphanet_98895	Orphanet_98896

Issued 11-01-2022 Modified 11-01-2022

FAIR Data Point



FAIR Data Point

RESOURCE TYPE: CATALOG

Duchenne Data Platform
Catalog : <https://w3id.org/duchenne-fdp/catalog/ea2a751b-34e5-49cc-ad40-14346ac30676>

Content [34c4efc4-c87d-4dd2-99c5-ea5cdcaa0ea8e](#)

Description:

accessRights: This metadata file has no restrictions
conformsTo: Catalog Profile
description: The Duchenne Data Platform (DDP) Catalog
hasVersion: 0.0.1
identifier: <https://w3id.org/duchenne-fdp/catalog/ea2a751b-34e5-49cc-ad40-14346ac30676>
issued: 2022-01-11T12:55:21.861322432Z
language: en
license: [creativecommons.org/publicdomain_zero_1.0_rdf](#)
modified: 2022-01-11T12:55:23.062339323Z
publisher: Stichting Patient Project Productions
metadataIdentifier: <https://w3id.org/duchenne-fdp/catalog/ea2a751b-34e5-49cc-ad40-14346ac30676>
metadataIssued: 2022-01-11T12:55:21.861322432Z
metadataModified: 2022-01-11T12:55:27.353761686Z

Associated Taxonomy:

[HP_0000708](#) [HP_0000750](#) [HP_0000819](#) [HP_0001263](#) [HP_0001270](#) [HP_0001639](#)
[HP_0002014](#) [HP_0002019](#) [HP_0002020](#) [HP_0002194](#) [HP_0002650](#) [HP_0002987](#)
[HP_0003044](#) [HP_0003270](#) [HP_0005997](#) [HP_0006380](#) [HP_0006957](#) [HP_0008366](#)
[HP_0008981](#) [HP_0009473](#) [HP_0011675](#) [HP_0020110](#) [HP_0030193](#) [HP_0032321](#)

FAIR Data Point

FAIR FAIR Data Point Log In

FAIR Data Points

Filter: All 164 Active 23 Inactive 33 Unreachable 76 Invalid 29 Unknown 3

Endpoint	Registration	Modification	Status
https://fdp.castoredc.com	01-02-2021, 11:57:53	23-02-2022, 07:00:02	ACTIVE
https://directory.bbmri-eric.eu/api/fdp	27-01-2021, 15:30:32	23-02-2022, 05:45:00	ACTIVE
https://covid19research.nl/api/fdp	27-01-2021, 15:39:43	23-02-2022, 05:00:05	ACTIVE
https://fdp.sdsc.edu	01-05-2020, 23:44:58	23-02-2022, 04:03:04	ACTIVE
https://zks-docker.ukl.uni-freiburg.de/fairdatapoint	24-01-2022, 13:39:27	23-02-2022, 01:13:22	ACTIVE
http://178.63.49.197:8050	15-02-2022, 17:09:30	22-02-2022, 17:09:30	ACTIVE
https://fdp.x-omics.nl	29-11-2021, 20:46:23	22-02-2022, 15:14:19	ACTIVE
https://twoc.fair-dtls.surf-hosted.nl	04-03-2021, 11:32:52	22-02-2022, 08:50:19	ACTIVE
https://w3id.org/duchenne-fdp	19-05-2021, 09:54:04	22-02-2022, 07:51:42	ACTIVE
https://home.fairdatapoint.org	20-01-2021, 21:56:42	21-02-2022, 21:25:51	ACTIVE

Linked Data

Publishing FAIR Data: An Exemplar Methodology Utilizing PHI-Base

Principios Linked Data

1. Usar URIs (Uniform Resource Identifier) para identificar entidades
2. Usar URIs que son accesibles mediante el protocolo HTTP(S), para que usuarios o agentes automáticos puedan acceder a las entidades
3. Cuando se acceda a la entidad, proveer datos sobre la entidad en formatos estándar y abiertos, como RDF (Resource Description Framework)
4. Añadir en los datos que publicamos en RDF enlaces a las URIs de otras entidades, de modo que un usuario o agente pueda navegar por la red de datos y descubrir más datos que también siguen los principios Linked Data

Linked Data: "Base de datos universal"

Utilizar maquinaria Web (URIs HTTP), para identificar y localizar entidades

Utilizar un modelo de datos común, tripleta RDF, para integrar datos en los que aparecen esa entidades

base de datos universal

Ventajas Linked Data

Descubrimiento e integración de datos

Programación de agentes que consuman los datos

Actualización de datos mediante enlaces

Consultas complejas

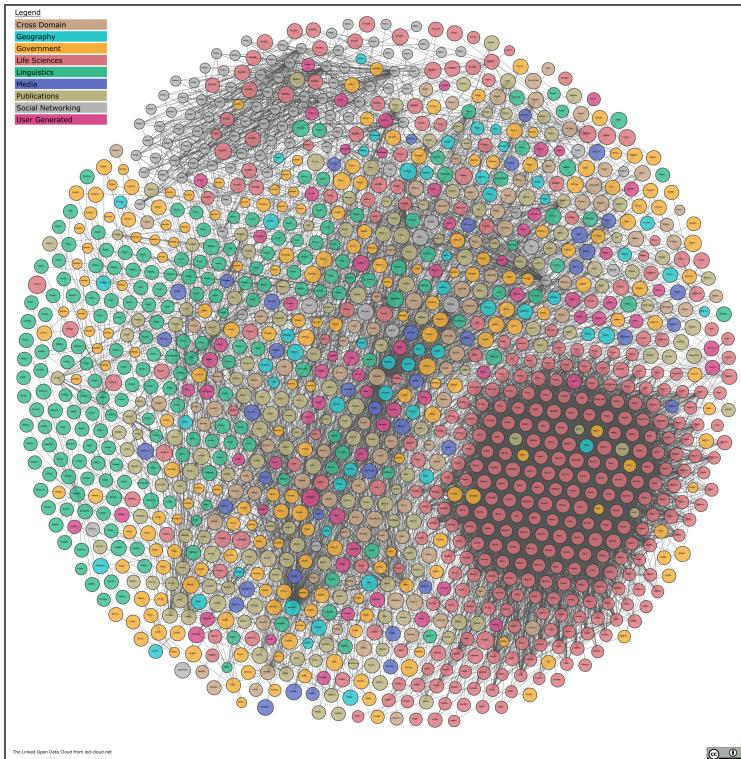
Ventajas Linked Data

Con Linked Data cualquiera puede publicar datos y enlazarlos a otros datos

El conjunto de datos abiertos publicados mediante Linked Data forma la «nube Linked Open Data»

Cada vez más instituciones públicas de todo el mundo usan Linked Data para publicar sus datos

Linked Open Data cloud



Linked Data: negociación de contenido

```
curl -L -H "Accept: text/html" "http://dbpedia.org/resource/Berlin"
```

```
| curl -L -H "Accept: application/rdf+xml" "http://dbpedia.org/resource/Berlin"
```

```
<owl:sameAs rdf:resource="http://eu.dbpedia.org/resource/Berlin" />
<owl:sameAs rdf:resource="http://linkedgeodata.org/triplify/node240109189" />
<owl:sameAs rdf:resource="http://sws.geonames.org/2958159/" />
```



```
curl -L -H "Accept: text/html" "http://sws.geonames.org/2950159/"
```

A screenshot of a web browser window. On the left is a map of a rural area with several locations marked. A callout box highlights a specific location with a blue marker and a white info box containing text. On the right is the raw HTML code for the page, which includes meta tags for character encoding, title, and viewport.

```
curl -L -H "Accept: application/rdf+xml" "http://sws.geonames.org/2950159/"
```

```
<@prefix country:><http://sws.geonames.org/country#>
<@prefix population:><http://sws.geonames.org/population#>
<@prefix pos:><http://sws.geonames.org/pos#>
<@prefix long:><http://sws.geonames.org/long#>
<@prefix geo:><http://sws.geonames.org/geo#>
<@prefix id:><http://sws.geonames.org/id#>
<@prefix feature:><http://sws.geonames.org/feature#>
<@prefix parentCountry:><http://sws.geonames.org/2921044#>
<@prefix parentADM1:><http://sws.geonames.org/2956157#>
<@prefix parentADM3:><http://sws.geonames.org/6547383#>
<@prefix parentADM4:><http://sws.geonames.org/6547357#>
<@prefix featureNearby:><http://sws.geonames.org/2956153#nearby,r>
```

Linked Data: negociación de contenido

```
curl -L -H "Accept: text/html" "http://dbpedia.org/resource/Berlin"
```

```
curl -L -H "Accept: application/rdf+xml" "http://dbpedia.org/resource/Berlin"
```

```
curl -L -H "Accept: text/html" "http://sws.geonames.org/2950159/"
```

```
curl -L -H "Accept: application/rdf+xml" "http://sws.geonames.org/2950159/"
```

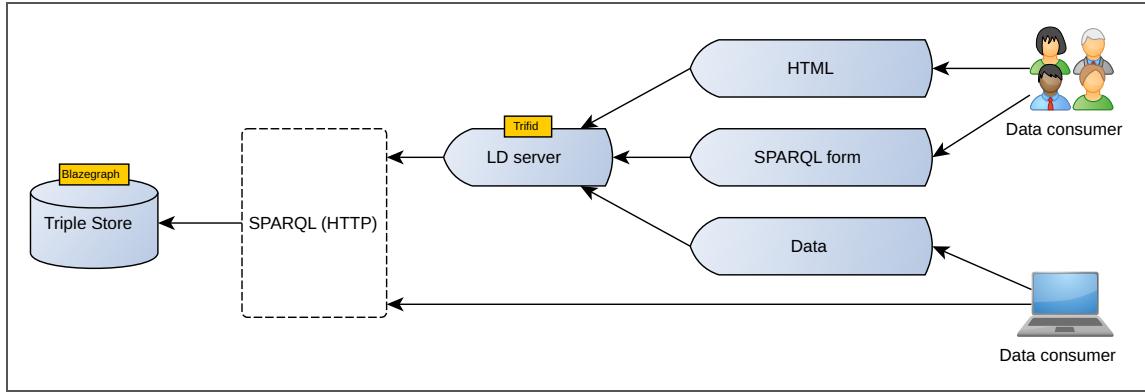
URIs/URLs en Linked Data

URI identifica a entidad; URLs localizan diferentes representaciones (RDF, HTML, ...) de la entidad

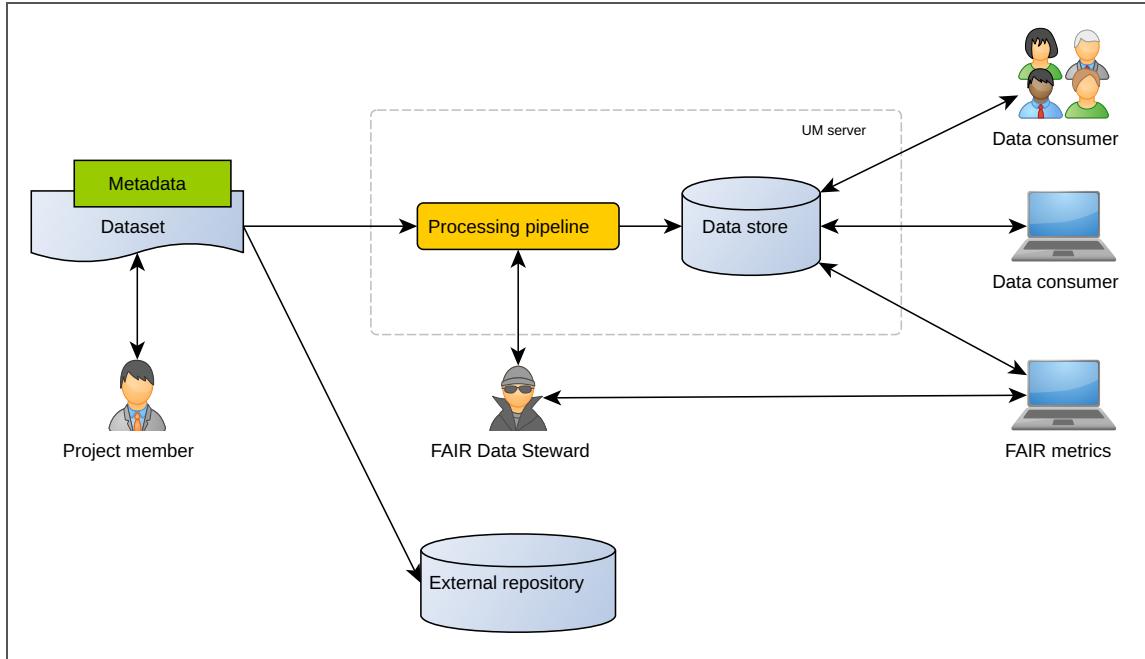
Descripción de la entidad (RDF, HTML, ...) ≠ entidad

HTTP URI dereferenciable: cuando se busca una URI, debería devolver una descripción adecuada del objeto que identifica esa URI

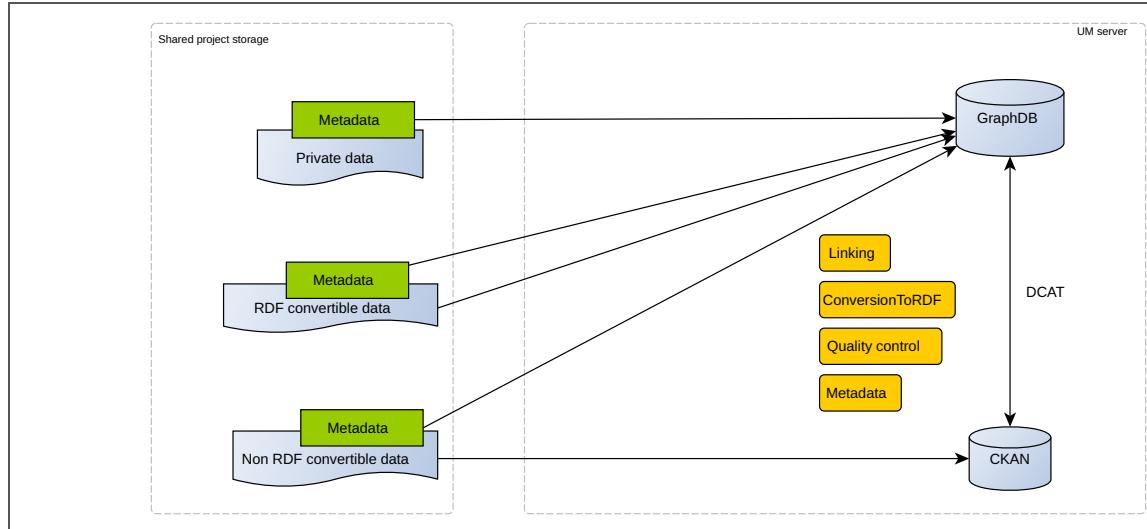
Publicar datos en Linked Data



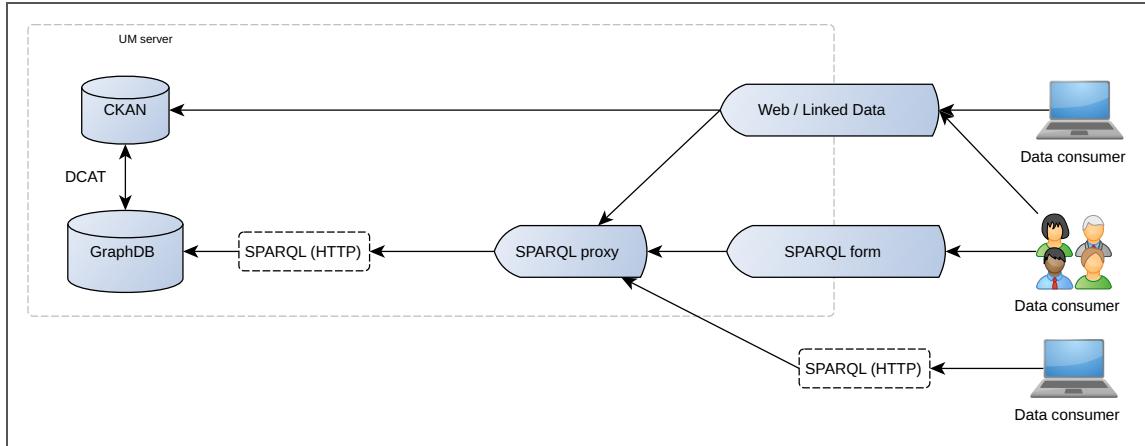
Publicar datos en Linked Data: SUPPORT4LHS



Publicar datos en Linked Data: SUPPORT4LHS



Publicar datos en Linked Data: SUPPORT4LHS



Ejemplo práctico

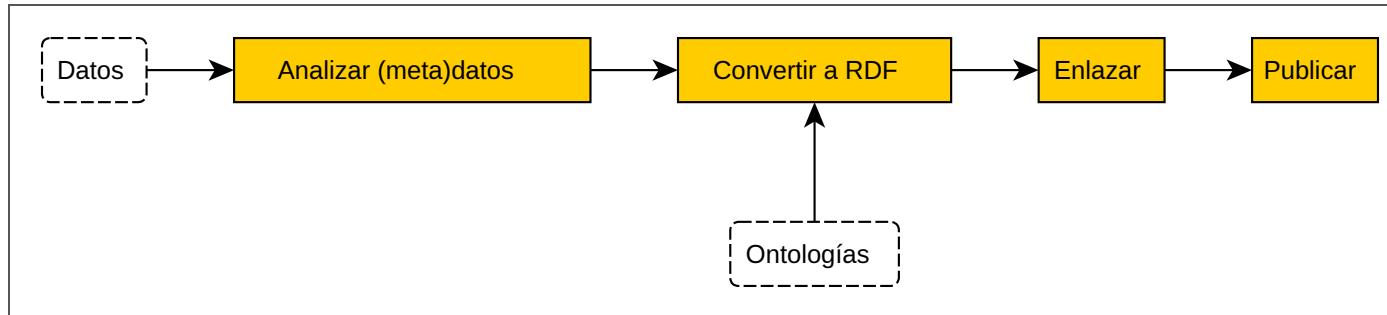
"FAIRificar" un dataset de ejemplo

Proceso vertical: intentar cubrir todos los pasos técnicos, sin entrar en detalles de contenidos

Ejemplos muy simples, nada realistas

A Generic Workflow for the Data FAIRification Process

566-Annikajacobsen-18.pdf



Datos de origen

[GenesUM.csv](#) (LinkedDataServer/data/)

Datos en RDF

[GenesUM.nt](#) (LinkedDataServer/data/)

[GenesUM.ng](#) (LinkedDataServer/data/)

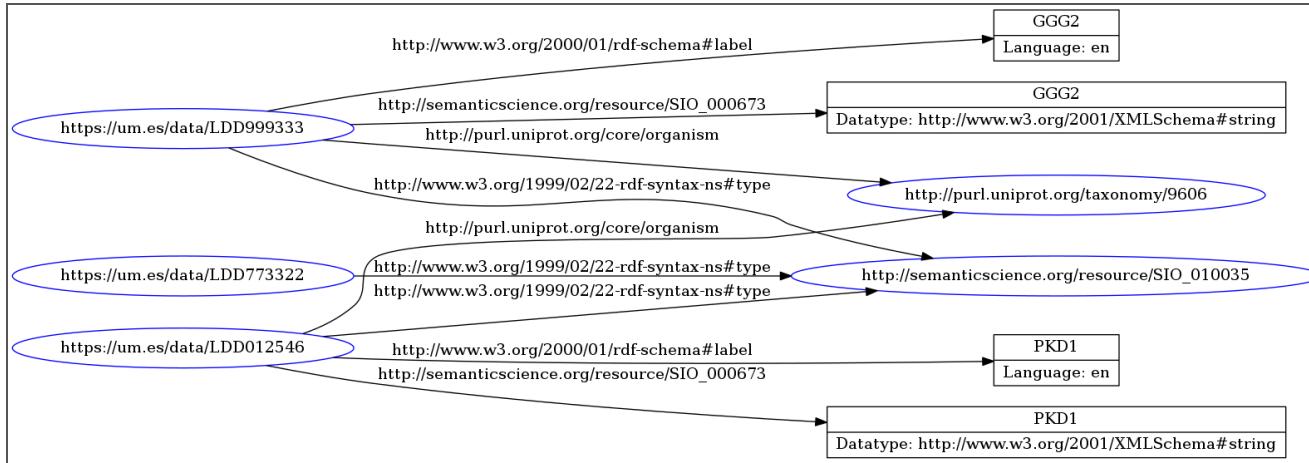
Conversión a RDF

[CSV2RDF.py](#)

Se basa en [RDFLib](#)

Otras herramientas posibles: [YARRRML](#) ([Google Enterprise Knowledge Graph Entity Reconciliation Service](#)), [TARQL](#), [OntoRefine](#), [Open Refine](#), [Eccenca CMEM](#), [Apache Any23](#), etc.

Conversión a RDF



Anotar con ontologías

Semantic Science Integrated Ontology (SIO) ([SIO_010035](#))

Uniprot Core Ontology ([9606](#))

Enlazar

A otras URIs

Manualmente, o con herramientas como [SILK](#)

Metadatos

Asignar una URI a nuestro dataset (F1):

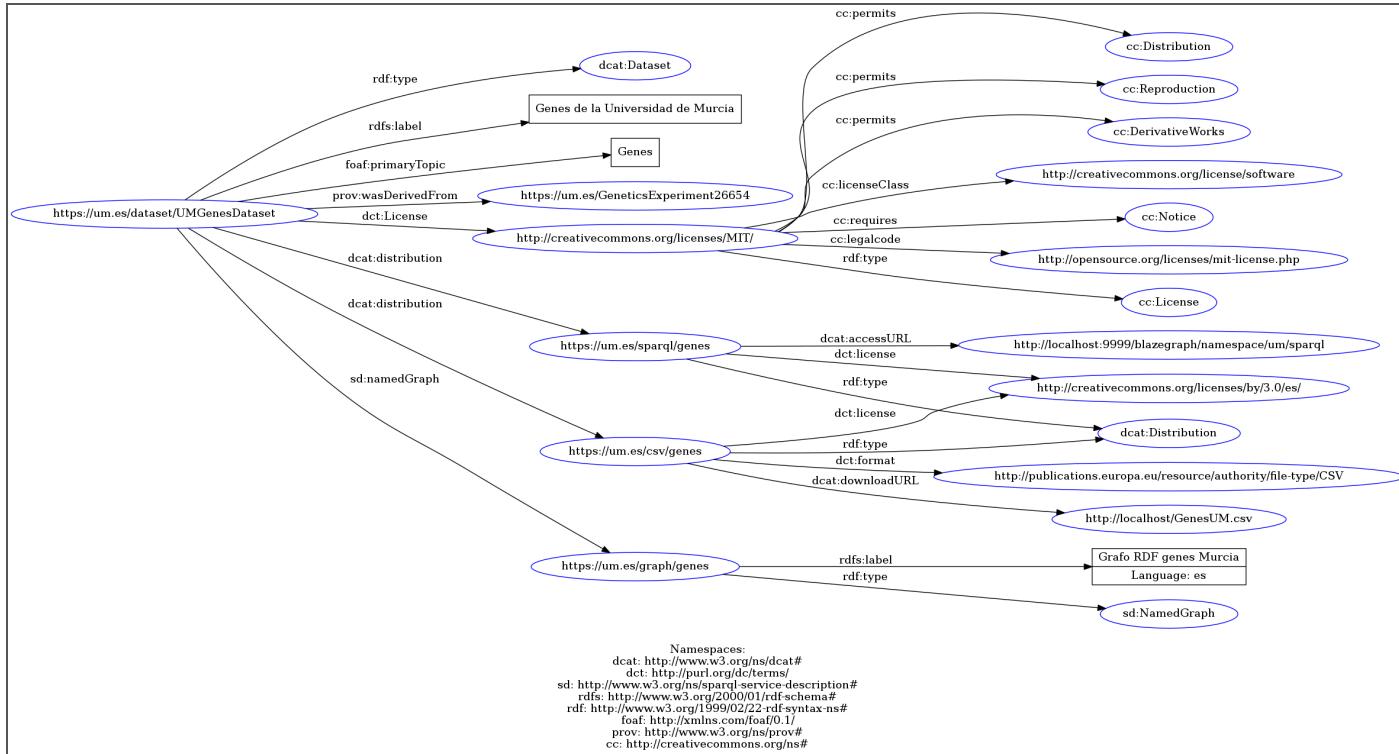
<https://um.es/dataset/UMGenesDataset>

Usar diferentes vocabularios como [DCAT](#), [VOID](#), [PROV](#), [FOAF](#), etc. para añadir metadatos

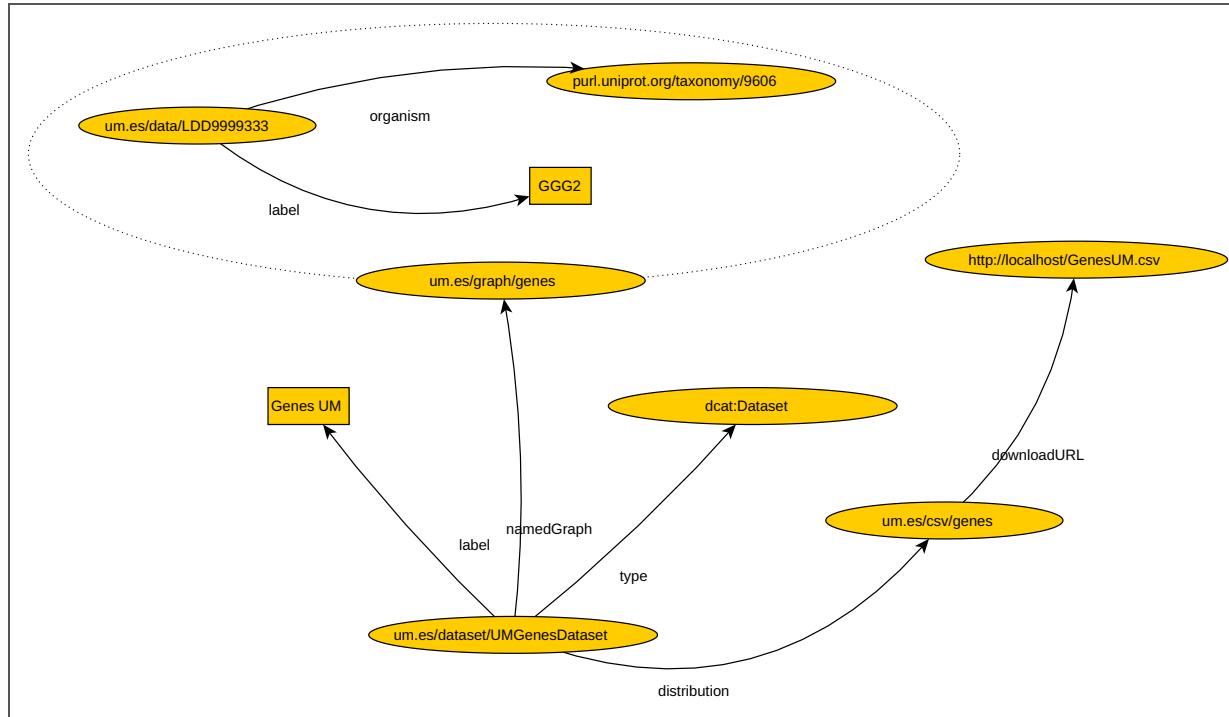
Named Graph: conjunto de triples que se identifica con una URI ([RDF W3C](#))

[MetadataGenes.ttl](#) (LinkedDataServer/data/)

Metadata



Datos/Metadatos



Datos/Metadatos

Datos: CSV (GenesUM.csv), RDF (GenesUM.nq)

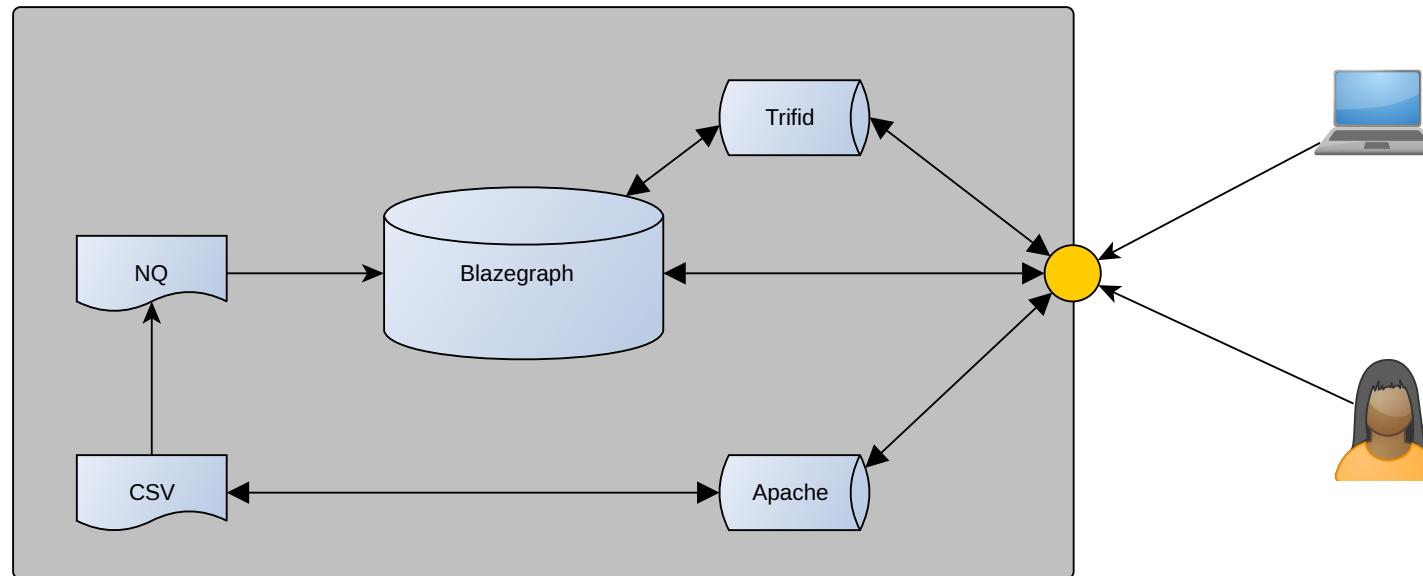
Metadatos: RDF (MetadataGenes.ttl)

Publicación

Datos: CSV (Apache), RDF (Blazegraph/Trifid)

Metadatos: RDF (Blazegraph/Trifid)

Publicación



Blazegraph, Trifid

LinkedDataServer/TrifidBlazegraph: \$ docker-compose up -d

```
version: "3"
services:
  linked_data_server:
    image: ghcr.io/zazuko/trifid
    ports:
      - "8080:8080"
    environment:
      SPARQL_ENDPOINT_URL: "http://sparql_endpoint:9999/blazegraph/namespace/um/sparql"
      #DATASET_BASE_URL: "https://um.es/data/"
      DATASET_BASE_URL: "http://fair/data/"

  sparql_endpoint:
    image: blazegraph
    ports:
      - "9999:9999"
```

Blazegraph

The screenshot shows the Blazegraph web interface with a red ribbon banner at the top right. The main navigation bar includes tabs for WELCOME, QUERY, UPDATE, EXPLORE, and NAMESPACES. The NAMESPACES tab is active, displaying the 'Namespaces' section. Below the title, there is a horizontal menu with links: kb, Use, Delete, Properties, Rebuild Full Text Index, Clone, and Service Description. A link to 'Download VOID description of all namespaces' is also present. The 'Create namespace' section follows, containing instructions and a configuration form. The form includes fields for 'Name' (set to 'um'), 'Mode' (set to 'quads'), and 'Inference' (disabled). It also has a 'Enable geospatial' checkbox and a 'Create namespace' button.

Namespaces

kb Use Delete Properties Rebuild Full Text Index Clone Service Description

[Download VOID description of all namespaces](#)

Create namespace

There are a number of features to enable. There's full documentation [here](#). You must select "Use" after A quick reference is below:

- o PropertyGraph: Select triples.
- o RDF + SPARQL with named graphs: Select quads mode.
- o Support for [Reification Done Right \(RDR\)](#): Select rdr mode.

Name: Mode: Inference: (Inference disabled -

Enable geospatial:

Trifid

Configurar Trifid para que haga consultas contra SPARQL endpoint de Blazegraph (Servicio sparql_endpoint):

http://sparql_endpoint:9999/blazegraph/namespace/um/sparql

[SPARQL Service Description](#)

[SPARQL Protocol](#)

Trifid

URIs:

- <https://um.es/data/>
- <http://IP:8080>

Apache

/var/www/html/

Resumen

¿Hemos conseguido implementar todos los principios FAIR?

¿En qué medida?

Proyecto a realizar

Reproducir en el servidor Google Cloud este proceso (Con documentación extra)

Incluir en Entregable de Explotación semántica de datos (Publicación de datos)