FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Una manera de hacer Europa

Bilbao, noviembre-2019

EF2.1-6. DOCUMENTO DE ANÁLISIS DE MÉTODOS FAIR

Contents

1. Introducción

Los Principios FAIR[1] proporcionan directrices para la publicación de recursos digitales tales como conjuntos de datos, códigos, flujos de trabajo y objetos de investigación, de manera que sean localizables, accesibles, interoperables y reutilizables (FAIR).

En 2016, se publicó en Scientific Data el artículo seminal sobre principios FAIR, titulado "*FAIR Guiding Principles for scientific data management and stewardship*" [1]. En tal artículo, los autores pretendieron proporcionar directrices para mejorar la búsqueda, la accesibilidad, la interoperabilidad y la reutilización de los activos digitales. Los principios ahí formulados enfatizan la capacidad de acción de las máquinas (es decir, la capacidad de los sistemas computacionales para encontrar, acceder, interoperar y reutilizar datos sin intervención humana o con una intervención mínima) porque los seres humanos dependen cada vez más del apoyo computacional para manejar los datos como resultado del aumento en el volumen, la complejidad y la velocidad de creación de los datos.

Los principios se refieren a tres tipos de entidades: datos (o cualquier objeto digital), metadatos (información sobre ese objeto digital) e infraestructura. Por ejemplo, el principio F4 define que tanto los metadatos como los datos se registren o indexen en un recurso encontrable (el componente de infraestructura). A continuación, resumimos los 4 principios FAIR.

1.1. FINDABLE (ENCONTRABLE/LOCALIZABLE)

El primer paso para (re)utilizar los datos es encontrarlos. Los metadatos y los datos deben ser fáciles de encontrar tanto para los seres humanos como para los ordenadores. Los metadatos legibles por máquinas son esenciales para el descubrimiento automático de conjuntos de datos y servicios, por lo que este es un componente esencial del proceso de FAIRification. A continuación, aparecen numerados los principios FAIR relativos a (F)IND:

- F1. A los (meta)datos se les asigna un identificador global único y persistente.
- F2. Los datos se describen con metadatos ricos (definidos por R1 a continuación)
- F3. Los metadatos incluyen clara y explícitamente el identificador de los datos que describen.

[1] https://doi.org/10.25504/FAIRsharing.WWI10U

- F4. Los (meta)datos se registran o indexan en un recurso de búsqueda

1.2. ACCESIBLE

Una vez que el usuario encuentra los datos requeridos, necesita saber cómo se puede acceder a ellos, posiblemente incluyendo autenticación y autorización.

- A1. Los (meta)datos son recuperables mediante su identificador utilizando un protocolo de comunicaciones normalizado.
    - A1.1 El protocolo es abierto, gratuito y de aplicación universal
    - A1.2 El protocolo permite un procedimiento de autenticación y autorización, cuando sea necesario
- A2. Los metadatos son accesibles, incluso cuando los datos ya no están disponibles.

1.3. INTEROPERABLE

Por lo general, los datos deben integrarse con otros datos. Además, los datos deben interoperar con aplicaciones o flujos de trabajo para su análisis, almacenamiento y procesamiento.

- I1. Los (meta)datos utilizan un lenguaje formal, accesible, compartido y ampliamente aplicable para la representación del conocimiento.
- I2. Los (meta)datos utilizan vocabularios que siguen los principios de FAIR
- I3. Los (meta)datos incluyen referencias cualificadas a otros (meta)datos

1.4. REUTILIZABLE

El objetivo final de FAIR es optimizar la reutilización de los datos. Para lograr esto, los metadatos y los datos deben estar bien descritos para que puedan ser replicados y/o combinados en diferentes entornos.

- R1. Los meta(datos) se describen ampliamente con una pluralidad de atributos precisos y relevantes
    - R1.1. Los (meta)datos se liberan con una licencia de uso de datos clara y accesible.

- R1.2. Los (meta)datos están asociados a una procedencia detallada
- R1.3. Los (meta)datos cumplen con los estándares de la comunidad relevantes para el dominio

1.5. APLICACIÓN DE LOS PRINCIPIOS FAIR EN HERCULES-ASIO

El objetivo principal del PT1 de este proyecto es crear la infraestructura ontológica que describa los datos que almacenará el SGI y que se concretará en la 'Red de Ontologías Hércules' (ROH). Este PT tiene también la tarea de sentar las bases y establecer los procedimientos, métricas y evaluación para alinear el proyecto con los principios FAIR (Findable, Accesible, Interoperable, Reusable) de cara tanto a la publicación de las ontologías como a la publicación de datos.

Como resultado, se implementará un sistema de comprobación automatizado del nivel FAIR cumplido por los recursos publicados durante el desarrollo del proyecto, que devolverá el nivel FAIR alcanzado, asociado a la versión del proyecto y la fecha en la que se ejecutó la comprobación. La solución resultante tendrá las siguientes propiedades.

- Ofrecerá datos que sean *Findable* (Encontrables) a través de un identificador persistente e incluyendo metadatos
- *Accessible* (Accesible) a través del protocolo universal HTTP
- Interoperable usando vocabularios ampliamente adoptados y
- Reusable, se publican usando licencias de uso que promocionen la reusabilidad, como por ejemplo Creative Commons 4.0 BY-SA.

Hoy en día, la mejor manera de publicar datos FAIR es hacerlo mediante Linked Data, teniendo especial cuidado de generar datos y metadatos de alta calidad, mejorando así la reusabilidad de los datos para máquinas (y, como consecuencia y en última instancia, para humanos).

El Sistema de Comprobación FAIRness Automatizado en HERCULES-ASIO será administrable desde el propio Backend a través de una interfaz de web privada. Al igual que el resto de componentes que desarrollaremos, tendrá una arquitectura SOA (Service Oriented Architecture) que permitirá su reutilización desde otros sistemas mediante llamadas a las funciones del API. También sería reutilizable en otros proyectos adaptando el código publicado.

Este entregable es el primero de cuatro que abordan la alineación de la solución propuesta con los principios FAIR:

- Análisis de métodos FAIR (Entregable: documento de análisis). Este es el documento presente.

  Entregable EF2-1.6: documento de análisis de métodos FAIR

- Definir y desarrollar un módulo de métricas FAIR para evaluación automatizada de los recursos (ontologías o datos)

  Entregable EF2-1.7: Métricas FAIR I- software y documentación de métricas

- Publicar los resultados de ejecutar las métricas FAIR

  Entregable EF2-1.8a: Métricas FAIR II - Resultados publicados en la Web de la evaluación con métricas FAIR, versión 1 (primer ciclo de desarrollo realizado - 80%)

  Entregable EF2-1.8b: Métricas FAIR II - Resultados publicados en la Web de la evaluación con métricas FAIR, versión 2.

En definitiva, proponemos la creación de un sistema automatizado de comprobación de métricas FAIR que obtendrá información acerca del cumplimiento de los principios FAIR. La definición precisa de las métricas se ha realizado en base a las métricas definidas en el proyecto FAIR Metrics [2], [3] y en métricas adicionales definidas por la UTE (pendientes de consideración).

Además de la obtención de las métricas FAIR descritas en el proyecto FAIR Metrics (https://github.com/FAIRMetrics/Metrics) proponemos que el sistema de comprobación automatizado genere otros indicadores o métricas acerca del cumplimiento de los principios FAIR, que serán especificadas tras haber primero implementado las 14 métricas formuladas por FAIRmetrics.

2. Conceptos FAIR

HERCULES-ASIO dará lugar a un grafo de conocimiento donde cada nodo representará un Objeto Digital (Digital Object) – ver Figura 1 – representado siguiendo un vocabulario y conjunto de formatos estándar, acompañado de un identificador persistente (PID), metadatos y documentación.

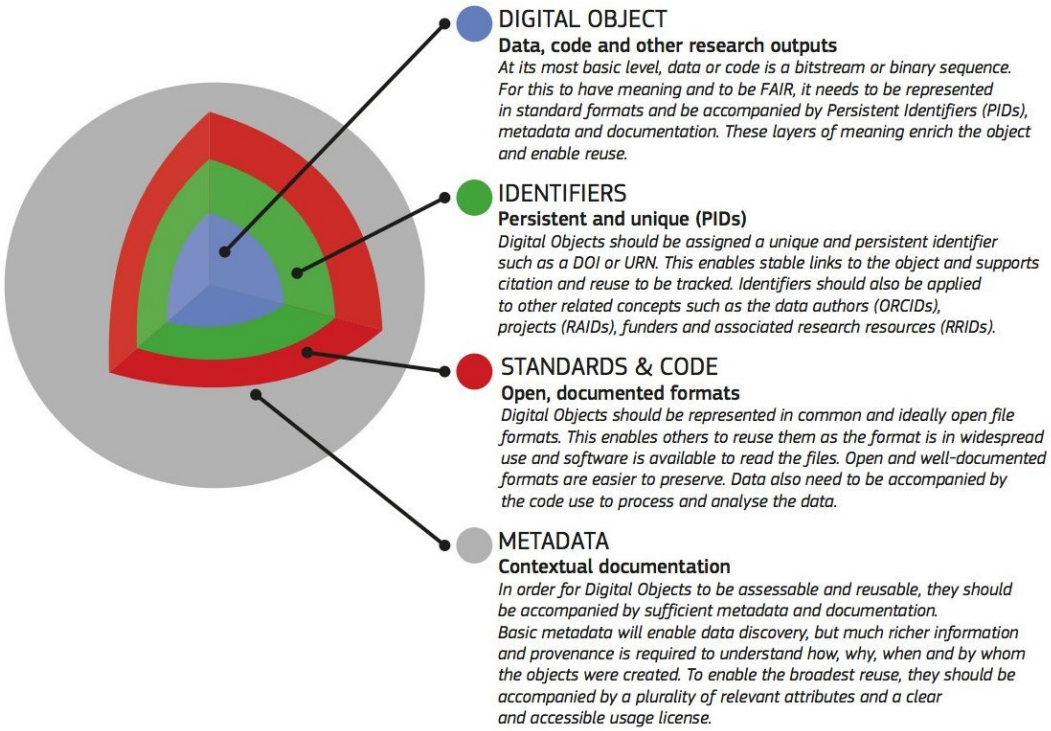

Figura 1. Modelo para objectos digitales FAIR (Findable, Accessible, Interoperable, Reusable) en HERCULES-ASIO [3].

Además, el SGI ASIO seguirá los principios Open Data y FAIR que garantizarán la búsqueda, acceso, interoperabilidad y reutilización de los datos de investigación de la Universidad de Murcia y del ecosistema de investigación estatal. Para ello, de acuerdo a los principios FAIR (ver Figura 2), se definirán las "políticas" que rijan la publicación de datos, un plan de gestión de datos que sirva de índice a los tipos de datasets que serán generados (este proyecto generará tal registro), usar convenciones para los identificadores asignados a los recursos modelados siguiendo estándares tanto respecto los identificadores como a los metadatos

asociados a las entidades modeladas[2]. Finalmente, el sistema HERCULES-ASIO debería poder facilitar la integración con terceros, es decir, no solamente ofrecer una interfaz administrativa para su uso dentro de la Universidad de Murcia sino que a través de una interfaz RESTful permita a terceras partes su interacción con HERCULES-ASIO. Asimismo, deberían respetarse los estándares ya existentes como OAI-PMH para garantizar su interoperabilidad con repositorios institucionales basados en software ampliamente adoptado por universidades internacionalmente como DSpace[3] o EPrints[4]. Por consiguiente, la UTE RIAM-DEUSTO debe hablar con la Universidad de Murcia para la integración de los datos que ya tienen en su Repositorio Institucional de la Universidad de Murcia[5].

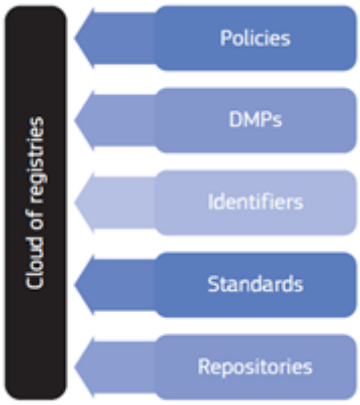

Figura 2. La realización de datos FAIR depende de los siguientes componentes esenciales: políticas, Planes de Gestión de Datos (DMP), identificadores y estándares [3].

[2] How OpenAIRE uses persistent identifiers for discovery, enrichment, and linking of ResearchResults, January 2019, https://www.slideshare.net/OpenAIRE_eu/how-openaire-uses-persistent-identifiers-for-discovery-enrichment-and-linking-of-research-results

[3] DSpace, https://duraspace.org/dspace/ , 2019.

[4] EPrints, https://www.eprints.org, 2019.

[5] https://digitum.um.es/

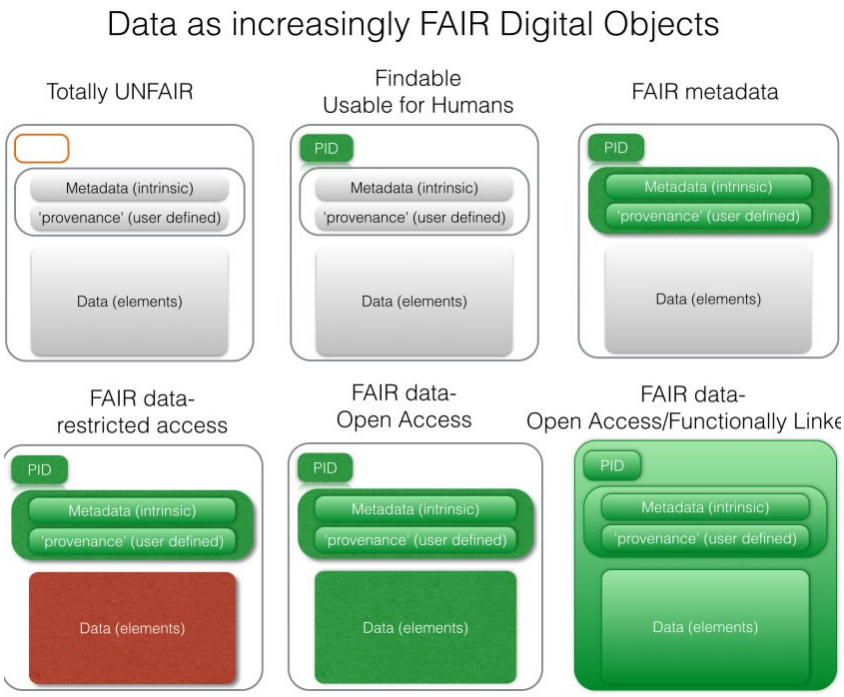

Figura 3. Enfoques para FAIRificación de recursos digitales [4].

El grado de cumplimiento de FAIRness por los objetos digitales puede seguir la clasificación ilustrada en la Figura 3, donde se muestran diferentes clasificaciones tanto del objeto digital como de los metadatos incluidos dentro de cada uno, que facilitan la usabilidad de los mismos tanto por personas como máquinas.

Una de las grandes preguntas en torno a los principios de FAIR ha sido cómo medir lo que llamamos su "FAIRness". La producción de un conjunto central de métricas semicuantitativas que tengan aplicabilidad universal para la evaluación de la FAIRness. Las métricas FAIR se pueden utilizar para cuantificar el grado en que un recurso digital es Findable, Accesible, Interoperable y Reutilizable. El trabajo propuesto en este entregable está basado íntegramente en lo especificado en el paper "*A design framework and exemplar metrics for FAIRness*" [5]. Tal trabajo propone 14 métricas, cada una de las cuales sigue la plantilla mostrada en la Figura 4.

FIELD	*DESCRIPTION*
Metric Identifier	FAIR Metrics should, themselves, be FAIR objects, and thus should have globally unique identifiers.
Metric Name	A human-readable name for the metric
To which principle does it apply?	Metrics should address only one sub-principle, since each FAIR principle is particular to one feature of a digital resource; metrics that address multiple principles are likely to be measuring multiple features, and those should be separated whenever possible.
What is being measured?	A precise description of the aspect of that digital resource that is going to be evaluated
Why should we measure it?	Describe why it is relevant to measure this aspect
What must be provided?	What information is required to make this measurement?
How do we measure it?	In what way will that information be evaluated?
What is a valid result?	What outcome represents "success" versus "failure"
For which digital resource(s) is this relevant?	If possible, a metric should apply to all digital resources; however, some metrics may be applicable only to a subset. In this case, it is necessary to specify the range of resources to which the metric is reasonably applicable.
Examples of their application across types of digital resource	Whenever possible, provide an existing example of success, and an example of failure.

Figura 4. Plantilla de métricas FAIR.

3. Alineación de HERCULES-ASIO con los principios FAIR

A continuación, especificamos las 14 métricas que serán puestas en marcha en el proyecto. Por cada métrica, adoptada del trabajo [1], exponemos detalles sobre cómo se abordará su implementación en HERCULES-ASIO (ver Tabla 1). Obsérvese que el documento tomado como referencia ha sido incluido como "Anexo I". Entendemos que el conjunto de métricas finales no se limitará a estas 14. Tal como indicamos en conclusiones, durante la elaboración de la herramienta para medición del FAIRness del ROH, podrán surgir nuevas métricas, algunas específicas al dominio de gestión de datos de investigación del proyecto HERCULES-ASIO.

Principio FAIR	Métrica/Explicación	Especificación	Verificación
Findable F1. (Meta)data are assigned a globally unique and persistent identifier	**FM-F1A. Identifier Uniqueness** The uniqueness of an identifier is a necessary condition to unambiguously refer that resource, and that resource alone. Otherwise, an identifier shared by multiple resources will confound efforts to describe that resource, or to use the ident to retrieve it.	An URL to a registered identifier scheme must be specified. An identifier scheme is valid if and only if it is described in a repository that can register and present such identifier schemes (e.g. fairsharing.org – a curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies [6]).	URL to a registered identifier scheme must be present: • A first version of this metric would focus on just checking a URL that resolves to a document. • A second version would indicate how to structure the data policy document with a section (similar to how the CC licenses now have a formal structure in RDF). • A third version would insist that that document and section is signed by an approved organization and made available in an appropriate repository.
¿Cómo es abordada en HERCULES-ASIO?	Tal como se reflejó en análisis de entidades efectuado en el "ANEXO I de la FASE I – Estudio de Viabilidad", para cada entidad se ha explorado el formato del identificador único a usar, estableciéndose mapeos entre el esquema de URIs uniforme y unívoco propuesto para Hércules e identificadores estándar que ya existen (ORCID, DOI). La factoría de URIs en HERCULES también jugará un importante rol aquí, asegurando que las IDs generados sean únicos y también persistentes. Esta métrica garantizará que las URIs de un recurso y sus metadatos puedan ser resueltas correctamente. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_unique_identifier.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.		
Findable F1. (Meta)data are assigned a globally unique and persistent identifier	**FM-F1B. Identifier persistence** Whether there is a policy that describes what the provider will do in the event an identifier scheme becomes deprecated.	Providers of digital resources must ensure that they have a policy to manage changes in their identifier scheme, with a special emphasis on maintaining/redirecting previously generated identifiers. They must provide a URL that resolves to a document containing the relevant policy.	Use an HTTP GET on URL provided. Present (a 200,202,203 or 206 HTTP response after resolving all and any prior redirects. e.g. 301 -> 302 -> 200 OK) or Absent (any other HTTP code). A first version of this metric would focus on just checking a URL that resolves to a document.
¿Cómo es abordada en ASIO?	Muy relacionado con métrica FM-F1A. El ID asignado no sólo tiene que ser único, sino que también apunte de manera permanente a un recurso incluso cuando cambie el identificador. Este es un requisito importante que la Factoría de URIs y		

		también el backend del SGI deben cumplir. La implementación inicial probará que la solución HERCULES-ASIO soporta el cambio de IDs de una entidad y que todavía se mantienen mapeos usando esquemas de URIs anteriormente puestos en marcha. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_identifier_persistence.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.	
Findable F2. Data are described with rich metadata	**FM-F2. Machine-readability of metadata** The availability of machine-readable metadata that describes a digital resource. This metric is intended to test the format of the metadata since machine readability of metadata makes it possible to optimize discovery. For instance, Web search engines suggest the use of structured metadata elements to optimize search. Thus, the machine-readability aspect can help people and machines and a digital resource of interest.	A URL to a document that contains machine-readable metadata for the digital resource must be provided. Furthermore, the file format must be specified.	HTTP GET on the metadata URL. A response of [a 200,202,203 or 206 HTTP response after resolving all and any prior redirects. e.g. 301 -> 302 -> 200 OK] indicates that there is indeed a document. The second URL should resolve to the record of a registered file format (e.g. DCAT, DICOM, schema.org etc.) in a registry like FAIRsharing. Possible valid results will be: a) Machine-readable or b) Machine-not-readable
¿Cómo es abordada en ASIO?		Para asegurarnos que ROH cumple está métrica es imprescindible que cada concepto representado ofrezca una URL apuntando a sus metadatos y que éstos estén en formatos estándar como DCAT – Data Catalogue Vocabulary[6] o Schema.org [7]. ROH será un catálogo de datasets de investigación, donde tanto el grafo de conocimiento per sé (`dcat:Catalog`) como cada dataset (`dcat:Dataset`) en particular serán descritos a través del vocabulario DCAT. Cada concepto definido en ROH debe proporcionar una propiedad `rdf:type` que vincule a cada entidad con una `rdfs:Class`. Además, la procedencia de los elementos será declarada explícitamente a través de propiedades de la ontología PROV-O. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_machine_readable_metadata.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.	

[6] https://www.w3.org/TR/vocab-dcat-2/

Findable F3. Metadata clearly and explicitly include the identifier of the data they describe	**FM-F3. Resource Identifier in Metadata** The discovery of a digital object should be possible from its metadata. This metric checks whether the metadata document contains the globally unique and persistent identifier for the digital resource.	The GUID[7] of the metadata and the GUID of the digital resource it describes, must be provided. For this to happen, the metadata must explicitly contain the identifier for the digital resource it describes, and this should be present in the form of a qualified reference, indicating some manner of "about" relationship, to distinguish this identifier from the numerous others that will be present in the metadata.	Parsing the metadata for the given digital resource GUID. If the GUID/UUID of the referred digital resource is present or absent. In addition, since many digital objects cannot be arbitrarily extended to include references to their metadata, in many cases, the only means to discover the metadata related to a digital object will be to search based on the GUID of the digital object itself.
¿Cómo es abordada en HERCULES-ASIO?	Íntimamente ligada a FM-F2, requiere que las conexiones entre recursos digitales y metadatos sean bidireccionales. En este caso, FM-F3 evalúa si desde los metadatos se puede llegar a los datos del recurso digital modelado. Por lo tanto, en ROH los metadatos de una entidad tienen que apuntar de vuelta necesariamente al recurso descrito incluyendo como valor la URI única y persistente del recurso digital. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_identifier_in_metadata.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.		
Findable F4. (Meta)data are registered or indexed in a searchable resource	**FM-F4. Indexed in a searchable resource** The degree to which the digital resource can be found using web-based search engines. Most people use a search engine to initiate a search for a particular digital resource of interest. If the resource or its metadata are not indexed by web search engines, then this would substantially diminish an individual's ability to find and reuse it.	The ability to discover a resource should be tested using a) its identifier, b) other text-based metadata. Taking as input the persistent identifier of the resource, perform search in the web and verify that pages pointing to the resource are returned.	We perform an HTTP GET on the URLs provided and attempt to find the persistent identifier in the page that is returned. A second step might include following each of the top N hits and examine the resulting documents for presence of the identifier. The result will be true when the persistent identifier was found in the search results.

[7] http://guid.one/guid

¿Cómo es abordada en HERCULES-ASIO?		Dada la URI de un recurso digital en ROH, deberemos extraer las URIs que apuntan a sus metadatos y realizar búsquedas con buscadores de Internet para explorar sus primeros resultados y determinar si se devuelven enlaces al GUID del recurso buscado y sus metadatos. Esta misma funcionalidad de búsqueda debería ser disponible a través de las herramientas de búsqueda de HERCULES-ASIO. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_searchable_index.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.	
Accessible **A1. (Meta)data are retrievable by their identifier using a standardised communications protocol** *A1.1 - the protocol is open, free, and universally implementable*	**FM-A1.1. Access Protocol** The nature and use limitations of the access protocol. Access to a resource may be limited by the specified communication protocol. In particular, we are worried about access to technical specifications and any costs associated with implementing the protocol. Protocols that are closed source or that have royalties associated with them could prevent users from being able to obtain the resource.	This metric should supply: • A URL to the description of the protocol • true/false as to whether the protocol is open source • true/false as to whether the protocol is (royalty) free	Do an HTTP get on the URL to see if it returns a valid document. Ideally, we would have a universal database of communication protocols from which we can check this URL. The HTTP GET on the URL should return a 200,202,203 or 206 HTTP response after resolving all and any prior redirects. e.g. 301 - 302 - 200 OK. The other two should return true/false ("true" is success)
¿Cómo es abordada en HERCULES-ASIO?		En ASIO, los identificadores utilizados serán resolubles a través de HTTP/s. Es decir, apuntan a recursos cuyos metadatos son devueltos a través de comandos HTTP (GET) en diferentes formatos atendiendo a la capacidad de negociación de contenidos del paradigma REST, que se basa en la cabecera `Content-Type`, pudiendo ser servidor en formatos estándar como RDF/XML, Turtle, RDFa o JSON-LD, dependiendo de la cabecera `Accept` de la petición HTTP. Uno de los componentes de ASIO es un servidor Linked Data que contestaría con datos en respuesta a una URI, en el formato indicado en la petición. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_access_protocol.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.	
Accessible **A1. (Meta)data are retrievable by their**	**FM-A1.2. Access authorization** Specification of a protocol to access restricted content. Not all content can be made available	The outcomes of this metric should be: • true/false concerning whether authorization is needed	Computational validation of the data provided. A valid answer contains a true or false for the first question.

identifier using a standardised communications protocol *A1.2 - the protocol allows for an authentication and authorization procedure, where necessary*	without restriction. For instance, access and distribution of personal health data may be restricted by law or by organizational policy. In such cases, it is important that the protocol by which such content can be accessed is fully specified.	• a URL that resolves to a description of the process to obtain access to restricted content.	If true, an HTTP GET on the URL provided should return a 200, 202, 203, or 206 HTTP Response after resolving all redirects.
¿Cómo es abordada en HERCULES-ASIO?	En HERCULES-ASIO, se usarán HTTPS para incrementar la seguridad en la comunicación, adoptándose OAuth para la autenticación y autorización de acceso a recursos. Esta métrica evaluará que sólo usuarios autorizados podrán acceder a los recursos digitales que requieren autorización. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_access_authorization.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.		
Accessible **A2 - metadata are accessible, even when the data are no longer available**	**FM-A2. Metadata Longevity** The existence of metadata even in the absence/removal of Data. Cross-references to data from third-party's FAIR data and metadata will naturally degrade over time and become “stale links". In such cases, it is important for FAIR providers to continue to provide descriptors of what the data was to assist in the continued interpretation of those third-party data. As per FAIR Principle F3, this meta-data remains discoverable, even in the absence of the data, because it contains an explicit reference to the IRI of the data.	This metric must verify the existence of an URL to a formal metadata longevity plan.	Resolve the URL checking: • Successful resolution • Returns a document that represents a plan or policy of some kind • Preferably certified (e.g. DSA)

¿Cómo es abordada en HERCULES-ASIO?	En el back-end de HERCULES-ASIO verificaremos que, aunque algunos recursos digitales dejen de existir, todavía sus metadatos sean recuperables. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_metadata_longevity.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.		
Interoperable **I1 - (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation**	**FM-I1. Use a Knowledge Representation Language** Use of a formal, accessible, shared, and broadly applicable language for knowledge representation. The unambiguous communication of knowledge and meaning (what symbols are, and how they relate to one another) necessitates the use of languages that are capable of representing these concepts in a machine-readable manner.	URL to the specification of the language • The language must have a BNF (or other specification language) • The URL resolves (accessible) • The document has an IANA media-type[8] (i.e. it is sufficiently widely accepted and shared that it has been registered) • The language can be arbitrarily extended (e.g. PDBml can be used to represent knowledge, but only about proteins)	BNF (or other?) found, Media-type of the document is registered in FAIRSharing. Future: FAIRSharing has tags to indicate constrained vs. extendable languages?
¿Cómo es abordada en HERCULES-ASIO?	En HERCULES-ASIO, los metadatos se expresarán según el lenguaje estándar de la web semántica Ontology Web Language, OWL. Se hará un uso extensivo de Ontologías ampliamente utilizadas por terceros (como por ejemplo FOAF, Dublin Core o schema.org) lo que garantizará la interoperabilidad entre nuestro grafo de conocimiento ROH y otros externos. Además, los metadatos serán exportados en diferentes serializaciones de RDF como RDF/XML, JSON-LD o Turtle. Revisaremos las ontologías reutilizadas a lo largo del tiempo y contrastaremos que ROH en una gran mayoría (>80%) está basado en ontologías conocidas y ampliamente probadas. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_knowledge_language.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.		

[8] https://www.iana.org/assignments/media-types/media-types.xhtml

Interoperable **I2 - (meta)data use vocabularies that follow FAIR principles**	**FM-I2. Use FAIR Vocabularies** The metadata values and qualified relations should themselves be FAIR, for example, terms from open, community-accepted vocabularies published in an appropriate knowledge-exchange format. Data, and the provenance descriptors of the data, should (where reasonable) use vocabularies and terminologies that are, themselves, FAIR.	IRIs representing the vocabularies used for (meta)data must be provided.	Resolve IRIs, check FAIRness of the returned document(s). Successful resolution; document is amenable to machine-parsing and identification of terms within it. It may be possible to use FAIRSharing to validate these vocabularies.
¿Cómo es abordada en HERCULES-ASIO?	En HERCULES-ASIO, se va a crear una red de ontologías (ROH) basada en ontologías ampliamente aceptadas. Tales ontologías van a ser enlazadas entre ellas para asegurar su interoperabilidad. Se respetarán las licencias de uso y reutilización de cada una de ellas, reconociendo la autoría inicial de las mismas a través de relaciones basadas en la ontología PROV-O. Esta métrica deberá verificar que los vocabularios adoptados para modelar conocimiento son FAIR por sí mismos. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_fair_vocabularies.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.		
Interoperable **I3 - (meta)data include qualified references to other (meta)data**	**FM-I3. Use Qualified References** Relationships within (meta)data, and between local and third-party data, have explicit and 'useful' semantic meaning. For Interoperability, the relationships within and between data must be more semantically rich than “is (somehow) related to". Numerous ontologies include richer relationships that can be used for this purpose, at various levels of domain-specificity. For example, the use of SKOS for terminologies (e.g. exact matches) references/relations point	Linksets (in the formal sense) representing part or all of your resource must be provided.	The linksets must have qualified references. At least one of the links must be in a different Web domain (or the equivalent of a different namespace for non-URI identifiers) • References are qualified • Qualities are beyond “ref" or “is related to” • One of the cross-references points outwards to a distinct Namespace

	outwards to other resources, owned by third-parties; this is one of the requirements for 5 star linked data.		
¿Cómo es abordada en HERCULES-ASIO?	En HERCULES-ASIO, siguiendo el enfoque de Linked Data nuestras ontologías y las instancias de las mismas estarán enlazadas a otras ontologías y recursos de grafos externos (Wikidata, DBpedia), respectivamente. Esta métrica evaluará que la mayoría de las relaciones sean específicas al dominio y no genéricas (ej. `related_to` frente a `child_of`), cuando enlacen las diferentes entidades modeladas en el grafo resultante. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_has_linkset.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.		
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*	**FM-R1.1. Accessible Usage License** The existence of a license document, for BOTH (independently) the data and its associated metadata, and the ability to retrieve those documents. A core aspect of data reusability is the ability to determine, unambiguously and with relative ease, the conditions under which you are allowed to reuse the (meta)data. Thus, FAIR data providers must make these terms openly available. This applies both to data (e.g. for the purpose of third-party integration with other data) and for metadata (e.g. for the purpose of third-party indexing or other administrative metrics)	The IRI of the license (e.g. its URL) for the data license and for the metadata license must be provided.	Resolve the IRI(s) using its associated resolution protocol. Valid result: A document containing the license information
¿Cómo es abordada en HERCULES-ASIO?	En ASIO, la red de ontologías Hércules (ROH) resultante tendrá una licencia Creative Commons 4.0 BY-SA. Se verificará que las ontologías conectadas en esta red dispongan todas ellas de una licencia compatible. Además, las ontologías importadas serán enriquecidas con descripciones semánticas, a través de las propiedades `dc:rights` y `dc:license`, del vocabulario Dublin Cored. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_accessible_license.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.		

Reusable **R1. Meta(data) are richly described with a plurality of accurate and relevant attributes** *R1.2 - (meta)data are associated with detailed provenance*	**FM-R1.2. Detailed Provenance** That there is provenance information associated with the data, covering at least two primary types of provenance information: • Who/what/When produced the data (i.e. for citation) • Why/How was the data produced (i.e. to understand context and relevance of the data). Reusability is not only a technical issue; data can be discovered, retrieved, and even be machine-readable, but still not be reusable in any rational way. Reusability goes beyond “can I reuse this data?" to other important questions such as “may I reuse this data?", “should I reuse this data", and “who should I credit if I decide to use it?"	Several IRIs - at least one of these points to one of the vocabularies used to describe citational provenance (e.g. dublin core). At least one points to one of the vocabularies (likely domain-specific) that is used to describe contextual provenance (e.g. EDAM)	We resolve the IRI according to their associated protocols. In the future, we may be able to cross-reference these with FAIRSharing to confirm that they are "standard", and perhaps even distinguish citation vs. domain specific IRI 1 should resolve to a recognized citation provenance standard such as Dublin Core. IRI 2 should resolve to some vocabulary that itself passes basic tests of FAIRness
¿Cómo es abordada en HERCULES-ASIO?	En HERCULES-ASIO, se hará uso de la ontología PROV-O (https://www.w3.org/TR/prov-o) para declarar explícitamente la procedencia de las ontologías y sus atributos y datos de instancia reutilizados. Esta métrica comprobará que se ha declarado provenance de cada vocabulario incorporado en ROH. Punto de partida de la implementación será [2], implementado tanto en Perl (i.e. `metric_detailed_provenance_A.pm`) como en Ruby. Realizaremos conversión a Python a partir de los scripts en Perl y proporcionaremos front-end web desde donde lanzar cada test.		
Reusable **R1. Meta(data) are richly described with a plurality of accurate and relevant attributes**	**FM-R1.3. Meets Community Standards** Certification, from a recognized body, of the resource meeting community standards. Various communities have recognized that maximizing the usability of their data requires them to adopt a set of guidelines for metadata (often in the form of “minimal information about. . . " models). Non-compliance with these	A certification saying that the resource is compliant should be provided.	Validate the electronic signature as coming from a community authority (e.g. a Verisign signature). Successful signature validation Such certification services may not exist, but this principle serves to encourage the community to create both the standard(s) and the verification services for those standards.

R1.3 - (meta)data meet domain-relevant community standards	standards will often render a dataset `reuseless' because critical information about its context or provenance is missing. However, adherence to community standards does more than just improve reusability of the data. The software used by the community for analysis and visualization often depends on the (meta)data having certain fields; thus, non-compliance with standards may result in the data being unreadable by its associated tools. As such, data should be (individually) certified as being compliant, likely through some automated process (e.g. submitting the data to the community's online validation service)		
¿Cómo es abordada en HERCULES-ASIO?	En HERCULES-ASIO, esta métrica es de difícil verificación automática, será evaluada por la revisión manual de la ontología resultante por expertos y la evaluación de la usabilidad de la misma por los usuarios de ROH. Se prepararán cuestionarios que evalúen que ROH cumple las convenciones del dominio de los sistemas de gestión de investigación, sigue estándares y además corresponde con una ontología fácilmente interoperable y usable.		

Tabla 1. Alineación de principios FAIR en HERCULES-ASIO.

4. Conclusión

Teniendo en cuanta las recomendaciones del grupo de expertos que definieron los principios FAIR, y nuestro trabajo analizando cómo aplicar principios FAIR en HERCULES-ASIO, se pueden sacar las siguientes conclusiones:

- Las métricas deben abordar la multidimensionalidad de los principios de FAIR y abarcar todos los tipos de objetos digitales.
- Las métricas universales pueden complementarse con métricas adicionales específicas de los recursos que reflejen las expectativas de comunidades particulares. En este caso el dominio de las universidades y la gestión de la investigación.
- Las métricas en sí mismas, y cualquier resultado derivado de su aplicación, deben ser FAIR. Por esa razón, adoptaremos la definición semántica de métricas que ya han efectuado los creadores de las métricas FAIR. Por ejemplo, https://purl.org/fair-metrics/FM_R1.3
- Los estándares abiertos en torno a las métricas deberían fomentar un ecosistema vibrante de herramientas de evaluación de la imparcialidad. Sin embargo, la capacidad de escalar las evaluaciones de FAIRness a miles de millones, si no a trillones, de objetos digitales diversos es fundamental para que se puedan aplicar diversos enfoques a la evaluación de FAIRness (por ejemplo, autoevaluación, grupos de trabajo, crowd-sourcing, automatización). En el seno de este trabajo garantizaremos la ejecución de métricas sobre muestras significativas del ROH para así validar que los principios FAIR se siguen cumpliendo en el tiempo.
- Las evaluaciones de imparcialidad deben mantenerse actualizadas, y todas las evaluaciones deben ser versionadas, tener un sello de tiempo y ser accesibles al público. Se creará un repositorio donde se ofrezcan reportes regulares sobre el seguimiento de los principios FAIR dentro de ROH.

- Las evaluaciones de imparcialidad, presentadas como una simple visualización, serán una poderosa modalidad para informar a los usuarios y guiar el trabajo de los productores de recursos digitales. Se ofrecerá la capacidad en la solución HERCULES-ASIO para visualizar el histórico de cumplimiento de métricas FAIR fácilmente visualizable a través de una tabla donde las filas serán las fechas en las que ejecutaron los tests y las columnas el grado de cumplimiento de cada métrica. Se usarán códigos de colores siguiendo la metáfora del semáforo.
- El proceso de evaluación, y la evaluación de la imparcialidad resultante, deben diseñarse y difundirse de manera que incentiven positivamente a los proveedores de recursos digitales; es decir, deben considerar que el proceso es justo e imparcial y, además, deben beneficiarse de estas evaluaciones y utilizarlas como una oportunidad para identificar áreas de mejora.
- La gobernanza de las métricas, y los mecanismos para evaluarlas, serán necesarios para permitir su evolución cuidadosa y abordar los desacuerdos válidos.

En la reimplementación en Python, refinamiento y extensión de las métricas identificadas dentro del proyecto HERCULES-ASIO se propone hacer uso de herramientas de procesado de texto como REGEX y SHACL [8] – lenguaje de restricciones para Grafos RDF que puede ser usado para describir la estructura de los datos - ya sea almacenados en RDF o JSON o formatos similares. SHACL proporciona un vocabulario RDF para clases, propiedades y restricciones de integridad casi arbitrarias que las instancias necesitan cumplir.

El resultado final será la generación de un test-suite que evalúe las 14 métricas descritas, así como otras adicionales, asociadas a los sistemas de gestión de investigación, que surjan como resultado de plasmar este modelo teórico de métricas en realidad. Se ofrecerá un front-end (interfaz de usuario) que facilite el lanzamiento del test suite y la generación de reportes que informen sobre el grado de cumplimiento de las métricas FAIR. Tales reportes serán archivados (históricos) para poder analizar el compromiso a lo largo del tiempo de ROH con los principios FAIR.

Bibliografía

[1] M. D. Wilkinson *et al.*, «The FAIR Guiding Principles for scientific data management and stewardship», *Scientific Data*, vol. 3, n.º 1, pp. 1-9, mar. 2016.
[2] Mark Wilkinson, Erik Schultes, Luiz Olavo Bonino, Susanna-Assunta Sansone, Peter Doorn, y Michel Dumontier, *FAIRMetrics/Metrics: FAIR Metrics, Evaluation results, and initial release of automated evaluator code*. Zenodo, 2018.
[3] P. O. of the E. Union, «Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data.», 26-nov-2018. [En línea]. Disponible en: https://op.europa.eu:443/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF. [Accedido: 23-nov-2019].
[4] F. Admin, «Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0», *FORCE11*, 10-sep-2014. [En línea]. Disponible en: https://www.force11.org/fairprinciples. [Accedido: 23-nov-2019].
[5] M. D. Wilkinson, S.-A. Sansone, E. Schultes, P. Doorn, L. O. B. da S. Santos, y M. Dumontier, «A design framework and exemplar metrics for FAIRness», *bioRxiv*, p. 225490, dic. 2017.
[6] «FAIRsharing». [En línea]. Disponible en: https://fairsharing.org/. [Accedido: 24-nov-2019].
[7] «Home - schema.org». [En línea]. Disponible en: https://schema.org/. [Accedido: 24-nov-2019].
[8] «Shapes Constraint Language (SHACL)». [En línea]. Disponible en: https://www.w3.org/TR/shacl/. [Accedido: 25-nov-2019].

Apéndice 1. Métricas publicadas por FAIRmetrics

FAIR Metrics ALL

Mark D. Wilkinson, Susanna-Assunta Sansone,
Erik Schultes, Peter Doorn,
Luiz Olavo Bonino da Silva Santos, Michel Dumontier

July 4, 2018

1

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
UNIVERSIDAD DE MURCIA
Unión Europea
FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)
Una manera de hacer Europa

FAIR Metrics ALL

Mark D. Wilkinson, Susanna-Assunta Sansone,
Erik Schultes, Peter Doorn,
Luiz Olavo Bonino da Silva Santos, Michel Dumontier

July 4, 2018

FIELD	*DESCRIPTION*
Metric Identifier	FM-F1A: https://purl.org/fair-metrics/FM_F1A
Metric Name	Identifier Uniqueness
To which principle does it apply?	F1
What is being measured?	Whether there is a scheme to uniquely identify the digital resource.
Why should we measure it?	The uniqueness of an identifier is a necessary condition to unambiguously refer that resource, and that resource alone. Otherwise, an identifier shared by multiple resources will confound efforts to describe that resource, or to use the identifier to retrieve it. Examples of identifier schemes include, but are not limited to URN, IRI, DOI, Handle, trustyURI, LSID, etc. For an in-depth understanding of the issues around identifiers, please see http://dx.plos.org/10.1371/journal.pbio.2001414
What must be provided?	URL to a registered identifier scheme.
How do we measure it?	An identifier scheme is valid if and only if it is described in a repository that can register and present such identifier schemes (e.g. fairsharing.org). Information about the identifier scheme must be presented with a machine-readable document containing the FM1 attribute with the URL to where the scheme is described. see specification for implementation.
What is a valid result?	Present or Absent
For which digital resource(s) is this relevant?	All

Examples of their application across types of digital resource	Ontology - Gene Ontology: http://www.ebi.ac.uk/miriam/main/ datatypes/MIR:00000022 - HISCO: [link] This resource has not described or registered their identifier scheme. A recommended course of action would be to XXX. Model/format - RDFS: https://fairsharing.org/bsgs000283 Repository - JWS Online: https://www.ebi.ac.uk/miriam/main /collections/MIR:00000130 - DANS EASY: Database - ArrayExpress: https://fairsharing.org/biodbcore-000305 -¿ FAIRsharing will implement the FAIR Metric specification to provide a machine-readable link to the MIRIAM repository (for life science content) API - smartAPI's API https://raw.githubusercontent.com/WebsmartAPI /smartAPI/master/docs/iodocs/smartapi.json –¿ the smartAPI repository will provide accessible specification of the identifier scheme that is embedded in that metadata document. Journal http://www.nature.com/developers/documentation/ metadata-resources/doi –¿ the web site will have to provide a machine-readable pointer to the official DOI specification.
Comments	A first version of this metric would focus on just checking a URL that resolves to a document. We can't verify that document. A second version would indicate how to structure the data policy document with a particular section (similar to how the CC licenses now have a formal structure in RDF). A third version would insist that that document and section is signed by an approved organization and made available in an appropriate repository.

FIELD	*DESCRIPTION*
Metric Identifier	FM-F1B: https://purl.org/fair-metrics/FM_F1B
Metric Name	Identifier persistence
To which principle does it apply?	F1
What is being measured?	Whether there is a policy that describes what the provider will do in the event an identifier scheme becomes deprecated.
Why should we measure it?	The change to an identifier scheme will have widespread implications for resource lookup, linking, and data sharing. Providers of digital resources must ensure that they have a policy to manage changes in their identifier scheme, with a specific emphasis on maintaining/redirecting previously generated identifiers.
What must be provided?	A URL that resolves to a document containing the relevant policy.
How do we measure it?	Use an HTTP GET on URL provided.
What is a valid result?	Present (a 200,202,203 or 206 HTTP response after resolving all and any prior redirects. e.g. 301 -¿ 302 -¿ 200 OK) or Absent (any other HTTP code)
For which digital resource(s) is this relevant?	All
Examples of their application across types of digital resource	for each of the 'canonical' data types, examples, if available. @todo FAIR principles (scholarly publication in Nature Scientific Data) https://www.doi.org/overview/DOI_article_ELIS3.pdf http://www.nature.com/developers/documentation/metadata-resources/doi/ FAIR Principles (computable representation): https://github.com/FAIRDataInitiative/FAIR-principles#fair-principles For DSA-certified repositories (example below of 3TUDatacentre at Delft) the identifier persistence policy is described in the self assessment: https://assessment.datasealofapproval.org/assessment_187/seal/pdf/ DOI Handbook - ensuring persistence: http://www.doi.org/doi_handbook/6_Policies.html#6.5}

Comments	A first version of this metric would focus on just checking a URL that resolves to a document. We can't verify that document.4 A second version would indicate how to structure the data policy document with a particular section (similar to how the CC licenses now have a formal structure in RDF). A third version would insist that that document and section is signed by an approved organization and made available in an appropriate repository.

FIELD	*DESCRIPTION*
Metric Identifier	FM-F2: https://purl.org/fair-metrics/FM_F2
Metric Name	Machine-readability of metadata
To which principle does it apply?	F2 - Data are described with rich metadata
What is being measured?	The availability of machine-readable metadata that describes a digital resource.
Why should we measure it?	This metric *does not* attempt to measure (or even define) "Richness" - this will be defined in a future Metric. This metric is intended to test the format of the metadata - machine readability of metadata makes it possible to optimize discovery. For instance, Web search engines suggest the use of particular structured metadata elements to optimize search. Thus, the machine-readability aspect can help people and machines find a digital resource of interest.
What must be provided?	A URL to a document that contains machine-readable metadata for the digital resource. Furthermore, the file format must be specified.
How do we measure it?	HTTP GET on the metadata URL. A response of [a 200,202,203 or 206 HTTP response after resolving all and any prior redirects. e.g. 301 -¿ 302 -¿ 200 OK] indicates that there is indeed a document. The second URL should resolve to the record of a registered file format (e.g. DCAT, DICOM, schema.org etc.) in a registry like FAIRsharing. Future ehnancements to FAIRSharing may include tags that indicate whether or not a given file format is generallyagreed to be machine-readable
What is a valid result?	Machine-readable or Machine-not-readable
For which digital resource(s) is this relevant?	All

Examples of their application across types of digital resource	This URL can resolve to: - A record in a metadata registry relevant to your digital object (e.g. FAIRsharing.org, FAIR Data Point, smartAPI editor) - Your metadata on an HTML web page using schema.org - A FAIR Accessor..... Semanticscience Integrated Ontology : http://semanticscience.org/ontology/sio.owl https://biosharing.org/bsg-s002686 Example of a DANS metadata-record of an archived dataset: https://easy.dans.knaw.nl/ui/datasets/id/easydataset:67859/tab/ 1 smartAPI's API metadata: https://raw.githubusercontent.com/WebsmartAPI/ smartAPI/master/docs/iodocs/smartapi.json5 Metadata record of a database: - GEO https://fairsharing.org/biodbcore-000441 Metadata record of a standard: - RDF https://fairsharing.org/bsg-s000559 Non-article Published Work - my Zenodo Deposit for polyA (https://doi.org/10.5281/zenodo.47641) - myExperiment Workflow

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
UNIVERSIDAD DE MURCIA
Unión Europea
FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)
Una manera de hacer Europa

FIELD	*DESCRIPTION*
Metric Identifier	FM-F3: https://purl.org/fair-metrics/FM_F3
Metric Name	Resource Identifier in Metadata
To which principle does it apply?	F3 - metadata clearly and explicitly include the identifier of the data it describes
What is being measured?	Whether the metadata document contains the globally unique and persistent identifier for the digital resource.
Why should we measure it?	The discovery of digital object should be possible from its metadata. For this to happen, the metadata must explicitly contain the identifier for the digital resource it describes, and this should be present in the form of a qualified reference, indicating some manner of ”about” relationship, to distinguish this identifier from the numerous others that will be present in the metadata. In addition, since many digital objects cannot be arbitrarily extended to include references to their metadata, in many cases the only means to discover the metadata related to a digital object will be to search based on the GUID of the digital object itself.
What must be provided?	The GUID of the metadata and the GUID of the digital resource it describes.
How do we measure it?	Parsing the metadata for the given digital resource GUID.
What is a valid result?	Present or absent
For which digital resource(s) is this relevant?	All
Examples of their application across types of digital resource	None
Comments	In practice there are issues related to the format of the metadata document that might make a simple string search impossible. For example, relative URLs in HTML and qnames in XML/RDF. We should engage in some community discussion about exactly how to execute this Metric.

FIELD	*DESCRIPTION*
Metric Identifier	FM-F4: https://purl.org/fair-metrics/FM_F4
Metric Name	Indexed in a searchable resource
To which principle does it apply?	F4 - (meta)data are registered or indexed in a searchable resource
What is being measured?	The degree to which the digital resource can be found using web-based search engines.
Why should we measure it?	Most people use a search engine to initiate a search for a particular digital resource of interest. If the resource or its metadata are not indexed by web search engines, then this would substantially diminish an individual's ability to find and reuse it. Thus, the ability to discover the resource should be tested using i) its identifier, ii) other text-based metadata.
What must be provided?	The persistent identifier of the resource and one or more URLs that give search results of different search engines.
How do we measure it?	We perform an HTTP GET on the URLs provided and attempt to to find the persistent identifier in the page that is returned. A second step might include following each of the top XX hits and examine the resulting documents for presence of the identifier.
What is a valid result?	true - the persistent identifier was found in the search results.
For which digital resource(s) is this relevant?	All

Examples of their application across types of digital resource	- my Zenodo Deposit for polyA (https://doi.org/10.5281/zenodo.47641) Test Query: 10.5281/zenodo.47641 orthology GOOGLE: Pass (#1 hit); BING: Fail (no hits); Yahoo: Fail (no hits); Baidu: Pass (#1 hit) Test Query: "protein domain orthology RNA Processing" Google: Pass (Hit #13); BING: Fail (not in top 40); Yahoo: Fail: (Not in top 40); Baidu: Pass (#1 Hit) - myExperiment Workflow (http://www.myexperiment.org/workflows/2969.html) Test Query: "workflow common identifiers EMC ontology" GOOGLE: Pass (#2 and #5 hit); BING: Fail (not in top 40, though OTHER workflows were found in top 10!); Yahoo: Fail (not in top 40, though other workflows found in top 10); Baidu: Pass (5/10 pages contained a link to the workflow, but the workflow itself was not discovered) - Jupyter notebook on GitHub (https://github.com/ VidhyasreeRamu/GlobalClimateChange/blob /master/GlobalWarmingAnalysis.ipynb) Test Query: "github python climate change earth surface temperature" Google: Fail (not in top 40; other similar Jupyter notebooks found in github); Bing: Fail (not in top 40... but MANY links to Microsoft Surface! LOL!); Yahoo: Fail (not in top 40); Baidu: Fail (not even a github hit in top 40!)
Comments	None

FIELD	*DESCRIPTION*
Metric Identifier	FM-A1.1: https://purl.org/fair-metrics/FM_A1.1
Metric Name	Access Protocol
To which principle does it apply?	A1.1 - the protocol is open, free, and universally implementable
What is being measured?	The nature and use limitations of the access protocol.
Why should we measure it?	Access to a resource may be limited by the specified communication protocol. In particular, we are worried about access to technical specifications and any costs associated with implementing the protocol. Protocols that are closed source or that have royalties associated with them could prevent users from being able to obtain the resource.

What must be provided?	i) A URL to the description of the protocol ii) true/false as to whether the protocol is open source iii) true/false as to whether the protocol is (royalty) free
How do we measure it?	Do an HTTP get on the URL to see if it returns a valid document. Ideally, we would have a universal database of communication protocols from which we can check this URL (this is now being created in FAIRSharing). We also check whether questions 2 and 3 are true or false.
What is a valid result?	The HTTP GET on the URL should return a 200,202,203 or 206 HTTP response after resolving all and any prior redirects. e.g. 301 - 302 - 200 OK. The other two should return true/false ("true" is success)
For which digital resource(s) is this relevant?	All
Examples of their application across types of digital resource	None
Comments	None

FIELD	*DESCRIPTION*
Metric Identifier	FM-A1.2: https://purl.org/fair-metrics/FM_A1.2
Metric Name	Access authorization
To which principle does it apply?	A1.2 - the protocol allows for an authentication and authorization procedure, where necessary
What is being measured?	Specification of a protocol to access restricted content.
Why should we measure it?	Not all content can be made available without restriction. For instance, access and distribution of personal health data may be restricted by law or by organizational policy. In such cases, it is important that the protocol by which such content can be accessed is fully specified. Ideally, electronic content can be obtained first by applying for access. Once the requester is formally authorized to access the content, they may receive it in some electronic means, for instance by obtaining an download URL, or through a more sophisticated transaction mechanism (e.g. authenticate, authorize), or by any other means. The goal should be to reduce the time it takes for valid requests to be fulfilled.

What must be provided?	i) true/false concerning whether authorization is needed ii) a URL that resolves to a description of the process to obtain access to restricted content.
How do we measure it?	computational validation of the data provided
What is a valid result?	a valid answer contains a true or false for the first question. if true, an HTTP GET on the URL provided should return a 200, 202, 203, or 206 HTTP Response after resolving all redirects.
For which digital resource(s) is this relevant?	All
Examples of their application across types of digital resource	None
Comments	None

FIELD	*DESCRIPTION*
Metric Identifier	FM-A2: https://purl.org/fair-metrics/FM_A2
Metric Name	Metadata Longevity
To which principle does it apply?	A2 - metadata are accessible, even when the data are no longer available
What is being measured?	The existence of metadata even in the absence/removal of data
Why should we measure it?	Cross-references to data from third-party's FAIR data and metadata will naturally degrade over time, and become "stale links". In such cases, it is important for FAIR providers to continue to provide descriptors of what the data was to assist in the continued interpretation of those third-party data. As per FAIR Principle F3, this metadata remains discoverable, even in the absence of the data, because it contains an explicit reference to the IRI of the data.
What must be provided?	URL to a formal metadata longevity plan
How do we measure it?	Resolve the URL

What is a valid result?	- Successful resolution - Returns a document that represents a plan or policy of some kind - Preferably certified (e.g. DSA)
For which digital resource(s) is this relevant?	All metadata
Examples of their application across types of digital resource	None
Comments	None

FIELD	*DESCRIPTION*
Metric Identifier	FM-I1: https://purl.org/fair-metrics/FM_I1
Metric Name	Use a Knowledge Representation Language
To which principle does it apply?	I1 - (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
What is being measured?	use of a formal, accessible, shared, and broadly applicable language for knowledge representation.
Why should we measure it?	The unambiguous communication of knowledge and meaning (what symbols are, and how they relate to one another) necessitates the use of languages that are capable of representing these concepts in a machine-readable manner.
What must be provided?	URL to the specification of the language
How do we measure it?	- The language must have a BNF (or other specification language) - The URL resolves (accessible) - The document has an IANA media-type (i.e. it is sufficiently widely-accepted and shared that it has been registered) - The language can be arbitrarily extended (e.g. PDBml can be used to represent knowledge, but only about proteins)
What is a valid result?	BNF (or other?) found, Media-type of the document is registered in FAIRSharing. Future: FAIRSharing has tags to indicate constrained vs. extendable languages?
For which digital resource(s) is this relevant?	All

Examples of their application across types of digital resource	None

Comments	michel: there must be a syntax and associated semantics for that language. This is sufficient mark: there needs to be some identity or denotation in the language; ('vanilla') xml and json are not FAIR, so should fail this test *** can you (i) identify elements and (ii) make statements about them, and iii) is there a formally defined interpretation for that -¿ HTML fails; PDF fails shared -¿ that there are many users of the language . acknowledged within your community -¿ hard to prove. . could we use google to query for your filetype (can't discriminate between different models) -¿ has a media type –¿ This SHOULD be stated as a IANA code [IANA-MT] standardization of at least this listing process is a good measure of "sharedness" broadly applicable . that the language is extensible to a domain of interest . you can define your own elements in accordance with the semantics of the language gff3 is not in the IANA list -¿ what steps would the community need to execute to be listed here? cases like GFF, PDB are not broadly applicable biopax -¿ is defined vnd.biopax.rdf+xml and built on rdf -¿ allows users to create new elements and relate them jpg -¿ widely used, registered, but primarily for image content pdf -¿ registered, enables users to create their own dictionary.

FIELD	*DESCRIPTION*
Metric Identifier	FM-I2: https://purl.org/fair-metrics/FM_I2
Metric Name	Use FAIR Vocabularies
To which principle does it apply?	I2 - (meta)data use vocabularies that follow FAIR principles

What is being measured?	The metadata values and qualified relations should themselves be FAIR, for example, terms from open, community-accepted vocabularies published in an appropriate knowledge-exchange format.
Why should we measure it?	It is not possible to unambiguously interpret metadata represented as simple keywords or other non-qualified symbols. For interoperability, it must be possible to identify data that can be integrated like-with-like. This requires that the data, and the provenance descriptors of the data, should (where reasonable) use vocabularies and terminologies that are, themselves, FAIR.
What must be provided?	IRIs representing the vocabularies used for (meta)data
How do we measure it?	Resolve IRIs, check FAIRness of the returned document(s).
What is a valid result?	Successful resolution; document is amenable to machineparsing and identification of terms within it. It may be possible to use FAIRSharing to validate these vocabularies.
For which digital resource(s) is this relevant?	All
Examples of their application across types of digital resource	None

Comments	michel: there must be a syntax and associated semantics for that language. This is sufficient mark: there needs to be some identity or denotation in the language; ('vanilla') xml and json are not FAIR, so should fail this test *** can you (i) identify elements and (ii) make statements about them, and iii) is there a formally defined interpretation for that -¿ HTML fails; PDF fails shared -¿ that there are many users of the language . acknowledged within your community -¿ hard to prove. . could we use google to query for your filetype (can't discriminate between different models) -¿ has a media type –¿ This SHOULD be stated as a IANA code [IANA-MT] standardization of at least this listing process is a good measure of "sharedness" broadly applicable . that the language is extensible to a domain of interest . you can define your own elements in accordance with the semantics of the language gff3 is not in the IANA list -¿ what steps would the community need to execute to be listed here? cases like GFF, PDB are not broadly applicable biopax -¿ is defined vnd.biopax.rdf+xml and built on rdf -¿ allows users to create new elements and relate them jpg -¿ widely used, registered, but primarily for image content pdf -¿ registered, enables users to create their own dictionary.

FIELD	*DESCRIPTION*
Metric Identifier	FM-I3: https://purl.org/fair-metrics/FM_I3
Metric Name	Use Qualified References
To which principle does it apply?	I3 - (meta)data include qualified references to other (meta)data
What is being measured?	Relationships within (meta)data, and between local and third-party data, have explicit and 'useful' semantic meaning

Why should we measure it?	One of the reasons that HTML is not suitable for machinereadable knowledge representation is that the hyperlinks between one document and another do not explain the nature of the relationship - it is "unqualified". For Interoperability, the relationships within and between data must be more semantically rich than "is (somehow) related to". Numerous ontologies include richer relationships that can be used for this purpose, at various levels of domainspecificity. For example, the use of skos for terminologies (e.g. exact matches), or the use of SIO for genomics (e.g. "has phenotype" for the relationship between a variant and its phenotypic consequences). The semantics of the relationship do not need to be "strong" - for example, "objectX wasFoundInTheSameBoxAs objectY" is an acceptable qualified reference Similarly, dbxrefs must be predicated with a meaningful relationship what is the nature of the cross-reference? Finally, data silos thwart interoperability. Thus, we should reasonably expect that some of the references/relations point outwards to other resources, owned by third-parties; this is one of the requirements for 5 star linked data.
What must be provided?	Linksets (in the formal sense) representing part or all of your resource

How do we measure it?	The linksets must have qualified references At least one of the links must be in a different Web domain (or the equivalent of a different namespace for non-URI identifiers)
What is a valid result?	- References are qualified - Qualities are beyond "Xref" or "is related to" - One of the cross-references points outwards to a distinct namespace
For which digital resource(s) is this relevant?	All
Examples of their application across types of digital resource	None
Comments	

FIELD	*DESCRIPTION*
Metric Identifier	FM-R1.1: https://purl.org/fair-metrics/FM_R1.1
Metric Name	Accessible Usage License
To which principle does it apply?	R1.1 - (meta)data are released with a clear and accessible data usage license
What is being measured?	The existence of a license document, for BOTH (independently) the data and its associated metadata, and the ability to retrieve those documents
Why should we measure it?	A core aspect of data reusability is the ability to determine, unambiguously and with relative ease, the conditions under which you are allowed to reuse the (meta)data. Thus, FAIR data providers must make these terms openly available. This applies both to data (e.g. for the purpose of third-party integration with other data) and for metadata (e.g. for the purpose of third-party indexing or other administrative metrics)
What must be provided?	IRI of the license (e.g. its URL) for the data license and for the metadata license
How do we measure it?	Resolve the IRI(s) using its associated resolution protocol
What is a valid result?	A document containing the license information
For which digital resource(s) is this relevant?	All
Examples of their application across types of digital resource	None
Comments	

FIELD	*DESCRIPTION*
Metric Identifier	FM-R1.2: https://purl.org/fair-metrics/FM_R1.2
Metric Name	Detailed Provenance
To which principle does it apply?	R1.2 - (meta)data are associated with detailed provenance

What is being measured?	That there is provenance information associated with the data, covering at least two primary types of provenance information: - Who/what/When produced the data (i.e. for citation) - Why/How was the data produced (i.e. to understand context and relevance of the data)
Why should we measure it?	Reusability is not only a technical issue; data can be discovered, retrieved, and even be machine-readable, but still not be reusable in any rational way. Reusability goes beyond "can I reuse this data?" to other important questions such as "may I reuse this data?", "should I reuse this data", and "who should I credit if I decide to use it?"
What must be provided?	Several IRIs - at least one of these points to one of the vocabularies used to describe citational provenance (e.g. dublin core). At least one points to one of the vocabularies (likely domain-specific) that is used to describe contextual provenance (e.g. EDAM)
How do we measure it?	We resolve the IRI according to their associated protocols. In the future, we may be able to cross-reference these with FAIRSharing to confirm that they are "standard", and perhaps even distinguish citation vs. domain specific.
What is a valid result?	IRI 1 should resolve to a recognized citation provenance standard such as Dublin Core. IRI 2 should resolve to some vocabulary that itself passes basic tests of FAIRness
For which digital resource(s) is this relevant?	All
Examples of their application across types of digital resource	None

Comments	Many data formats have fields specifically for Provenance information. -¿ could fairsharing curate these 4 fields? for every format and vocabulary? Some formats do not have these fields. For example, although gff can have arbitrary headers, the standard itself does not provide specific fields to capture detailed provenance. It therefore would

FIELD	*DESCRIPTION*
Metric Identifier	FM-R1.3: https://purl.org/fair-metrics/FM_R1.3
Metric Name	Meets Community Standards
To which principle does it apply?	R1.3 - (meta)data meet domain-relevant community standards
What is being measured?	Certification, from a recognized body, of the resource meeting community standards.
Why should we measure it?	Various communities have recognized that maximizing the usability of their data requires them to adopt a set of guidelines for metadata (often in the form of "minimal information about..." models). Non-compliance with these standards will often render a dataset 'reuseless' because critical information about its context or provenance is missing. However, adherence to community standards does more than just improve reusability of the data. The software used by the community for analysis and visualization often depends on the (meta)data having certain fields; thus, noncompliance with standards may result in the data being unreadable by its associated tools. As such, data should be (individually) certified as being compliant, likely through some automated process (e.g. submitting the data to the community's online validation service)
What must be provided?	A certification saying that the resource is compliant
How do we measure it?	Validate the electronic signature as coming from a community authority (e.g. a verisign signature)
What is a valid result?	Successful signature validation
For which digital resource(s) is this relevant?	All
Examples of their application across types of digital resource	None

Comments	Such certification services may not exist, but this principle serves to encourage the community to create both the standard(s) and the verification services for those standards. A potentially useful side-effect of this is that it might provide an opportunity for content-verification - e.g. the certification service provides a hash of the data, which can be used to validate that it has not been edited at a later date.

FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Una manera de hacer Europa