Document describing metrics for performing analyzes involving data repositories.

Description analysis

## Findable

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | F1 | | |
| **Summary** | F1: (Meta)data are assigned a globally unique and persistent identifier. | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º The (Meta) data must have some URI (Uniform Resource Identifier), such as:   * Compressed URI (CURIE) <prefix>: <local ID> * URIs (Prioritize what contains an identifier generated by the repository in the style <http URI pattern> <Local ID> - e.i.: https://www.uniprot.org/uniprot/P98161) * URLs (In some cases, however, they must necessarily contain the Local ID)   2º The repository must contain some standardization in the design of its local identifiers, in order to guarantee that they are globally unique.  3º Analyze if the identifiers are persistent. For that, the persistence identifiers must be consulted:   * identifiers.org * https://github.com/perma-id/w3id.org * http://re3data.org/ (You must search the repository and thus analyze whether it has ID standardization and persistence) * And others… | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Unique Identifier | Test whether the metadata resource has a unique identifier. | Only test if the URI is either a compressed URI or a normal URI |
| Identifier Persistence | Metric to test whether the metadata resource's unique identifier is likely to be persistent. The known scheme is registered with FAIRSharing (https://fairsharing.org/standards/?q=&selected\_facets=type\_exact:identifier%20schema). For URLs that do not follow a scheme in FAIRSharing, we tested known URL persistence schemes (purl, oclc, fdlp, purlz, w3id, ark). | If it is CURIE, see if it is in FAIRSharing.  If URI, test your persistence on:  PURL(<https://archive.org/services/purl/>),  OCLC (<http://bibpurl.oclc.org/maint/display.html>),  FDLP (Link not found)  PURLZ (Same as PURL),  W3ID (<https://github.com/perma-id/w3id.org>), ARK (<https://n2t.net/e/pub/naan_registry.txt> e <https://n2t.net/e/n2t_full_prefixes.yaml>) |
| Data Identifier Persistence | Metric to test whether the unique identifier for the data resource is likely to be persistent. The known scheme is registered with FAIRSharing (https://fairsharing.org/standards/?q=&selected\_facets=type\_exact:identifier%20schema). For URLs that do not follow a scheme in FAIRSharing, we tested known URL persistence schemes (purl, oclc, fdlp, purlz, w3id, ark). | If it is CURIE, see if it is in FAIRSharing.  If URI, test your persistence on:  PURL(<https://archive.org/services/purl/>),  OCLC (<http://bibpurl.oclc.org/maint/display.html>),  FDLP (Link not found),  PURLZ (Same as PURL),  W3ID (<https://github.com/perma-id/w3id.org>), ARK (<https://n2t.net/e/pub/naan_registry.txt> e <https://n2t.net/e/n2t_full_prefixes.yaml>) |

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | F2 | | |
| **Summary** | Data are described with rich metadata. | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º Metadata must have informative descriptions of context, quality and condition, or data characteristics.  2º It must be possible to find the data based on the information contained in the metadata, even without the identifier.  3º Intrinsic data (metadata captured automatically), contextual (metadata about the context the data was collected / generated), ontologies, taxonomies, controlled dictionaries, use of FAIR Data Point (FDP), make the metadata richer. If you follow FDP, take into account the specification at https://github.com/FAIRDataTeam/FAIRDataPoint-Spec/blob/master/spec.md | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Structured Metadata | Tests whether a machine is capable of finding structured metadata. They can be (for example) RDFa, embedded json, json-ld or structured metadata negotiated by content, such as RDF Turtle. | Python script based on the Extruct library that returns any of these formats is used.  Structured key / value metadata should be returned.  Structured metadata of the Linked Data type must be returned. |
| Grounded  Metadata | Tests whether a machine is capable of finding grounded metadata. that is, metadata terms that are in a resolvable namespace, where the resolution leads to a definition of the meaning of the term. Examples include JSON-LD, embedded schema or any form of RDF. This test currently excludes XML. | Some kind of Linked Data must be found.  Python script based on the Extruct library that returns any of these formats is used. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | F3 | | |
| **Summary** | Metadata clearly and explicitly include the identifier of the data they describe. | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º Analyse if the link to the data is available within the metadata, in case the data is present in another URI. If there is a link check if use persistent identifier.  2º Identify some standardized connection between metadata and data. | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Data Identifier Explicitly in Metadata | Metric to test whether metadata contains the data's unique identifier. This is done by looking for a variety of properties, including foaf: primaryTopic, schema: mainEntity, schema: distribution, sio: is-about and iao: is-about. The codeRepository scheme is used for software versions. | Applying the Extruct and Web Page Inspector script and analyzing (manually) if have some type of predefined property. |
| Metadata Identifier Explicitly in Metadata | Metric for testing whether the metadata contains the unique identifier for the metadata itself. This is done using a variety of scraping tools, including DOI metadata resolution, the use of the Python tool 'extruct' and other ... | Apply the Extruct and Web Page Inspector script and analyze (manually) if have some type of predefined property. From the result, if positive, analyze if it has unique IDs, CURIE, DOIs and others. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | F4 | | |
| **Summary** | (Meta)data are registered or indexed in a searchable resource. | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º From the URI and/or local ID search in some web-based search engines, in this case Google. | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Searchable in Major Search Engine | Test whether a machine is able to discover the resource by searching, using Google | Done manually. |

## Accessible

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | A1 | | |
| **Summary** | (Meta)data are retrievable by their identifier using a standardised communications protocol. | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º There should be no barriers to proprietary software or protocols for accessing (meta) data from their identifiers | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Don't have FM Gen2 | Don't have FM Gen2 | Don't have FM Gen2 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | A1.1 | | |
| **Summary** | The protocol is open, free, and universally implemented. | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º Analyse if there is a possibility that the data will be private, this in the entire repository.  2º Try to download (meta)data and see if there are any restrictions.  3º Analyse if there are means of accessing (meta)data via FTP or other means to facilitate Machine-Readable. | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Uses Open Free Protocol for Data Retrieval | The data can be recovered by an open and free protocol. Tests the data GUID for resolution protocol. Currently passes InChI keys, DOIs, identifiers and URLs. | Using the Extruct script together with the Web Page Inspector to analyse if any URI, CURIE is returned. |
| Uses Open Free Protocol for Metadata Retrieval | Metadata can be retrieved using an open and free protocol. Tests the metadata GUID for resolution protocol. Currently passes InChI keys, DOIs, identifiers and URLs. | Using the Extruct script together with the Web Page Inspector to analyse if any URI, CURIE is returned. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | A1.2 | | |
| **Summary** | The protocol allows for an authentication and authorisation procedure, where necessary. | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º Analyze if there is a possibility that the data will be private, this in the entire repository. | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Data Authentication and Authorization | Test the discovered data GUID for the ability to implement authentication and authorization in the resolution protocol. Currently passes InChI keys, DOIs, identifiers and URLs. Also searches the metadata for the Dublin Core 'accessRights' property, which can point to a document that describes the data access process. | The extruct script with web page inspector is used and if any URI or CURIE is returned. It is also analysed if there are any properties of type acessRights. |
| Metadata Authentication and Authorization | Tests the metadata GUID for the ability to implement authentication and authorization in its resolution protocol. Currently passes InChI keys, DOIs, identifiers and URLs. | TThe extruct script with web page inspector is used to see if any URI or CURIE is returned. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | A2 | | |
| **Summary** | Metadata are accessible, even when the data are no longer avaible. | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º Analyse if there is a possibility of (meta)data being changed.  2º If there is to see if there is any type of versioning of the (meta)data.  3º Try to access old versions (old IDs) to see if it is possible to find metadata.  4º Analyse if there is metadata referencing updated (meta)data. | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Metadata Persistence | Metric for testing whether metadata contains a persistence policy, explicitly identified by a key persistencePolicy (in hash data) or by a http://www.w3.org/2000/10/swap/pim/doc#persistencePolicy predicate in linked data. | The extruct script with web page inspector is used to see if there are any of the requested identifiers. |

## Interoperable

|  |  |  |  |
| --- | --- | --- | --- |
| **Principle** | I1 | | |
| **Summary** | (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º Analyse if there are common standardised controlled vocabularies, ontologies, thesaurus (containing persistent and globally unique identifiers) and if they use global identifiers.  2º Analyse if there is any standardization of (meta)data to guarantee interoperability (RDF, JSON-LD, DAML-OIL, OWL) | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Metadata Knowledge Representation Language (WEAK) | Maturity indicator for testing whether metadata uses a formal language widely applicable to knowledge representation. This particular test has a broad view of what defines a 'knowledge representation language'; in this assessment, anything that can be represented as structured data will be accepted. | Use a extruct script with web page inspector to see if there is any type of structured data and if meets this requirement. |
| Metadata Knowledge Representation Language (STRONG) | Maturity indicator for testing whether metadata uses a formal language widely applicable to knowledge representation. This particular test has a broad view of what defines a 'knowledge representation language'; in this evaluation, a knowledge representation language is interpreted as one in which the terms are semantically based on ontologies. Any form of RDF will pass this test (including RDF that is automatically extracted by third-party analyzers, such as Apache Tika). | An extruct script with web page inspector is used and if finds any CURIE check if it is in FAIRSharing. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | I2 | | |
| **Summary** | (Meta)data use vocabularies that follow FAIR principles. | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º Analyse if the documents for standardization of (meta) data comply with the FAIR Principles.  2º Analyse if there is any similarity or use FAIR Data Point. | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Metadata Uses Fair Vocabularies (WEAK) | Maturity indicator to test whether linked data metadata uses resolved terms. | An extruct script with web page inspector is used to analyse if any type of linked data occurs. |
| Metadata Uses Fair Vocabularies (STRONG) | Maturity indicator to test if the linked data metadata uses resolvable terms for linked data (FAIR). | The extruct script with web page inspector is used and if any type of structured data found meets this requirement.  Not done with Apache Tika |

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | I3 | | |
| **Summary** | (Meta)data include qualified references to other (meta)data | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º Analyse how the metadata refers to the data and others (meta)data.  2º Analyse if there is any kind of qualification among the references contained in the (meta)data, as it can be found at http://www.uniprot.org/uniprot/C8V1L6.rdf | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Metadata Contains Qualified Outward References | Test whether metadata is externally linked to third-party resources. It only tests metadata that can be represented as linked data. | Extruct script and web page inspector are used. It is analyzed if there is a link with external resources. |

## Reusable

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | R1 | | |
| **Summary** | (Meta)data are richly described with a plurality of accurate and relevant attributes | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º Because it is very similar to F2, it is also subjective about what it means in the (goal) given to be richly described. The more informative descriptions it contains the better. | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Don't have FM Gen2 | Don't have FM Gen2 | Don't have FM Gen2 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | R1.1 | | |
| **Summary** | (Meta)data are released with a clear and accessible data usage license. | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º It is analysed if there is any type of license present in the registration of the (meta)data or contained in the repository guidelines (e.g. Creative Commons).  2º It is analysed, in Re3Data, if the repository has any use license.  3º Check if there is data protection, ethical issues or conditions of reuse. | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Metadata Includes License (WEAK) | Maturity indicator for testing whether metadata contains an explicit license pointer. This 'Poor' test will use a case-insensitive regular expression and scan key/value style metadata as well as linked data metadata. Tests: license attributes xhtml, dvia, dcterms, cc, data.gov.au and Schema in linked data and validates the value of these properties. | Extruct script and web page inspector are used. It is analysed if there is any direction or referral to an access license. |
| Metadata Includes License (STRONG) | Maturity indicator to test whether the linked data metadata contains an explicit license pointer. Tests: license attributes xhtml, dvia, dcterms, cc, data.gov.au and Schema in linked data and validates the value of these properties. | Extruct script and web page inspector are used. It is analysed if there is an occurrence for an explicit use license pointer. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | R1.2 | | |
| **Summary** | (Meta)data are associated with detailed provenance | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º Analyse if there is any quote about provenance in the record of the (meta)data or in the repository as a whole.  2º Analyse if there is an incidence of prospective and/or retrospective origins. | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Don't have FM Gen2 | Don't have FM Gen2 | Don't have FM Gen2 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Description** | | |
| **Principle** | R1.3 | | |
| **Summary** | (Meta)data meet domain-relevant Community standards. | | |
| **How should it be analyzed?**  **(FAIR PRINCIPLES)** | 1º It is necessary to analyse whether the (meta) data are considered relevant by the community. How to accomplish this ?? | | |
| **How should it be analyzed?**  **(FAIR METRICS GEN2)** | **FM Gen2** | **Description** | **How is it being done?** |
| Don't have FM Gen2 | Don't have FM Gen2 | Don't have FM Gen2 |

Melhorias (Princípios FAIR e Métricas)

## Findable

### Human

#### F1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | All records have standardized identifiers (documentation is required), are globally unique (CURIE or URI) and are persistent (must be contained in some persistence repository or guarantee their persistence). |
| Average |  | 2 | All records have standardized identifiers (documentation required), are globally unique (CURIE or URI). They are not persistent. |
| Poor |  | 1 | All records are globally unique (CURIE or URI) OR are standardized. They are not persistent. |

#### F2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | Metadata has informative descriptions of context, quality and condition, or data characteristics. It must be possible to find the data from the metadata (even without an identifier). It must contain intrinsic, contextual data (ontologies, taxonomies, controlled dictionaries). Metadata is standardized. |
| Average |  | 2 | Metadata is not complete, some type of metadata is missing (context, quality, or data characteristic). It is possible to find the data from the metadata. Metadata is not fully standardized. |
| Poor |  | 1 | Metadata is not complete and / or is not standardized. No essential metadata is found. |

#### F3

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | There is a link to the data within the metadata, in case the data is present in another URI. An identification is established about the connection between metadata and data and it is standardized. Can use languages like RDF meet this criterion, being based on ontologies. |
| Average |  | 2 | The data and / or data link is found within the metadata, there is no definition of a standardization between the connection between data and metadata. |
| Poor |  | 1 | It is not possible to locate the data and / or data link within the metadata, it can even be found but it is not subjective. |

#### F4

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | The (meta) data is searchable, from its unique ID, in a known site indexer. The indexer must have additional information and not just a link to the (meta) data (the repository must allow the execution of crawlers to achieve this objective). |
| Average |  | 2 | The (meta) data is searchable, from its unique ID, in an indexer of known sites. There is no additional information in the indexer (the use of crawlers was not allowed). |
| Poor |  | 1 | The (meta) data is not searchable, from its unique ID, in a known site indexer. |

### Machine

#### FM-Gen2-F1.1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | Metadata has a unique identifier. |
| Average |  | 2 |  |
| Poor |  | 1 | Metadata does not have a unique identifier. |

#### FM-Gen2-F1.2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | The unique identifier for the metadata resource is not persistent. |
| Average |  | 2 |  |
| Poor |  | 1 | The unique identifier for the metadata resource is not persistent. |

#### FM-Gen2-F1.3

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | The unique identifier for the data resource is not persistent. |
| Average |  | 2 |  |
| Poor |  | 1 | The unique identifier for the data resource is not persistent. |

#### FM-Gen2-F2.1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | There is structured metadata. They can be (for example) RDFa, embedded json, json-ld or structured metadata negotiated by content, such as RDF Turtle. |
| Average |  | 2 |  |
| Poor |  | 1 | There is no structured metadata. |

#### FM-Gen2-F2.2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | Grounded metadata is found. Examples include JSON-LD, embedded schema or any form of RDF. |
| Average |  | 2 |  |
| Poor |  | 1 | No grounded metadata is found. |

#### FM-Gen2-F3.1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | The metadata contains the unique identifier for the data. foaf: primaryTopic, schema: mainEntity, schema: distribution, sio: is-about and iao: is-about, are examples to be found. The codeRepository scheme is used for software versions. |
| Average |  | 2 |  |
| Poor |  | 1 | There is no unique identifier for the data. |

#### FM-Gen2-F3.2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | Metadata contains the unique identifier for the metadata itself. Result of scraping tools, including the resolution of DOI metadata, the use of the Python tool 'extruct' and other ... |
| Average |  | 2 |  |
| Poor |  | 1 | Metadata does not contain a unique identifier for the metadata itself. |

#### FM-Gen2-F4

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | It is able to discover the resource by search, using a site indexer. |
| Average |  | 2 |  |
| Poor |  | 1 | It is not able to discover the resource by search, a using site indexer. |

## Accessible

### Human

#### A1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | There are no barriers to proprietary software or protocols to access (meta)data from their identifiers. |
| Average |  | 2 | There are barriers to proprietary software or protocols to access (meta)data from their identifiers, but they can be used by all users. |
| Poor |  | 1 | There are barriers to proprietary software or protocols to access (meta)data from their identifiers, however the users can’t find/use the softwares. |

#### A1.1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | The (meta)data is opened and it is possible to download it without authorization or user authentication restrictions. Use of other forms of access such as FTP are different. |
| Average |  | 2 | The (meta)data is opened and it is possible to download it, however restrictions on the download are found, such as proprietary software. |
| Poor |  | 1 | The (meta)data is opened and it is possible to download it, however it is not possible to download the data for any number of reasons, such as the use of proprietary software. |

#### A1.2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | Authentication and authorization steps are required to access the data. The description for meeting these phases is well documented with criteria for authentication and authorization. |
| Average |  | 2 | Authentication and authorization steps are required to access the data. There is no documentation to meet these phases. |
| Poor |  | 1 | There must be steps authentication and authorization for access to data, however it fails. |

#### A2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | The (meta)data can be changed. There is some type of versioning of (meta)data. IDs must be persistent. It is standardized and documented, such as the use of persistence policies. |
| Average |  | 2 | The (meta)data can be changed. There is some type of versioning of (meta)data. IDs must not be persistent. It is not standardized and documented. |
| Poor |  | 1 | The (meta)data can be changed. There is no versioning of the (meta)data. IDs must not be persistent. It is not standardized and documented. |

### Machine

#### FM-Gen2-A1.1.1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | The data can be recovered by an open and free protocol. Results such as InChI keys, DOIs, identifiers and URLs are accepted. |
| Average |  | 2 |  |
| Poor |  | 1 | The data cannot be recovered by an open and free protocol. |

#### FM-Gen2-A1.1.2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | Metadata can be retrieved using an open and free protocol. Results such as InChI keys, DOIs, identifiers and URLs are accepted. |
| Average |  | 2 |  |
| Poor |  | 1 | Metadata cannot be retrieved using an open and free protocol. |

#### FM-Gen2-A1.2.1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | The discovered data GUID is capable of implementing authentication and authorization in its resolution protocol. Results such as InChI keys, DOIs, identifiers and URLs are accepted. He also searches the metadata for the Dublin Core 'accessRights' property, which can point to a document that describes the data access process. |
| Average |  | 2 |  |
| Poor |  | 1 | The discovered data GUID is unable to implement authentication and authorization in its resolution protocol. |

#### FM-Gen2-A1.2.2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | The discovered metadata GUID is capable of implementing authentication and authorization in its resolution protocol. Results such as InChI keys, DOIs, identifiers and URLs are accepted. He also searches the metadata for the Dublin Core 'accessRights' property, which can point to a document that describes the data access process. |
| Average |  | 2 |  |
| Poor |  | 1 | The discovered metadata GUID is unable to implement authentication and authorization in its resolution protocol. |

#### FM-Gen2-A2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | Metadata contains a persistence policy, explicitly identified by a key persistencePolicy (in hashed data) or by a http://www.w3.org/2000/10/swap/pim/doc#persistencePolicy predicate in linked data. |
| Average |  | 2 |  |
| Poor |  | 1 | There is no policy of explicit persistence for a key. |

## Interoperable

### Human

#### I1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | There are common controlled vocabularies, ontologies, thesaurus (containing persistent and globally unique identifiers). There is standardization of (meta) data that guarantees interoperability (RDF, JSON-LD, DAML-OIL, OWL) |
| Average |  | 2 | There are common controlled vocabularies, ontologies and / or thesaurus (containing persistent and globally unique identifiers). There is no standardization of (meta) data that guarantees interoperability (RDF, JSON-LD, DAML-OIL, OWL) |
| Poor |  | 1 | There are no common controlled vocabularies, ontologies and / or thesaurus (containing persistent and globally unique identifiers). There is no standardization of (meta) data that guarantees interoperability (RDF, JSON-LD, DAML-OIL, OWL) |

#### I2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | The controlled vocabulary is used to describe datasets that are well documented and solvable using persistent and globally unique identifiers and this document is easily usable and accessible by everyone who uses the dataset. |
| Average |  | 2 | The controlled vocabulary used to describe datasets are not as well documented and solvable using persistent and globally unique identifiers and / or this document is not so easily usable and accessible by everyone who uses the dataset or may not yet exist. |
| Poor |  | 1 | There is no definition of controlled vocabulary to describe datasets using persistent and globally unique identifiers and / or this document is not usable and accessible by everyone who uses the dataset or does not exist. |

#### I3

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | (meta)data have qualifications among them, very well established, grounded and represented through knowledge representation languages (such as RDF) and ontologies. |
| Average |  | 2 | There is no qualification among the (meta)data, however there is the use of knowledge representation languages (such as RDF) |
| Poor |  | 1 | There is no qualification between the relationships of (meta) data and there is no use of knowledge representation languages (such as RDF). |

### Machine

#### FM-Gen2-I1.1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | Metadata uses a formal language widely applicable to knowledge representation. Anything that can be represented as structured data will be accepted. |
| Average |  | 2 |  |
| Poor |  | 1 | Metadata does not use a formal language widely applicable to knowledge representation. |

#### FM-Gen2-I1.2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | Metadata uses a formal language widely applicable to knowledge representation. In this assessment, a knowledge representation language is interpreted as one in which the terms are semantically based on ontologies. Any form of RDF will pass this test (including RDF that is automatically extracted by third-party analyzers, such as Apache Tika). |
| Average |  | 2 |  |
| Poor |  | 1 | Metadata does not use a formal language widely applicable to knowledge representation. |

#### FM-Gen2-I2.1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | Linked data metadata uses resolved terms. This only tests if they resolve, and FAIR data does not resolve, so it is a somewhat Poor test. |
| Average |  | 2 |  |
| Poor |  | 1 | Linked data metadata does not use resolved terms. |

#### FM-Gen2-I2.2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | Linked data metadata uses resolved terms for linked data (FAIR). |
| Average |  | 2 |  |
| Poor |  | 1 | Linked data metadata does not use resolved linked data (FAIR) terms. |

#### FM-Gen2-I3

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | Metadata links externally to third-party resources. It only tests metadata that can be represented as linked data. |
| Average |  | 2 |  |
| Poor |  | 1 | Metadata does not link externally to third-party resources |

#### 

## Reusable

### Human

#### R1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | (Meta)data is very well described with a plurality of attributes. There must be standardization and documentation of the data. |
| Average |  | 2 | (Meta)data is not so well described with a plurality of attributes. There may be some standardization and documentation of the data. |
| Poor |  | 1 | (Meta)data not described with plurality of attributes. There is no standardization and documentation of the data. |

#### R1.1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | It makes use of very well defined and documented licenses of use and that it is possible for the user to find them both in the registry and in the repository. |
| Average |  | 2 | It makes use of use licenses, but they are not so clear and defined. They may not be easily accessible to the user. |
| Poor |  | 1 | There is no specification of use licenses for both the repository and the registry. |

#### R1.2

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | There is specification of the prospective and retrospective provenance, it is very detailed and substantiated, it is present in the (meta)data. |
| Average |  | 2 | There is no specification of provenance by means of documents, but it is possible to find in the (meta)data provenance subjectively. |
| Poor |  | 1 | There is no documentation on provenance, there are (meta)data linked to provenance but they are not correct and/or are not complete. |

#### R1.3

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | There is confirmation and certification from the community about the (meta)data represented in the repository. |
| Average |  | 2 | The (meta)data is not certified by the community, but represents (meta)data consolidated by communities in the environment. |
| Poor |  | 1 | The (meta)data are not certified nor represent (meta)data consolidated by the middle communities. |

### Machine

#### FM-Gen2-R1.1.1

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | The metadata contains an explicit pointer to the license. This 'Poor' test will use a case-insensitive regular expression and scan key / value style metadata as well as linked data metadata. Tests: license attributes xhtml, dvia, dcterms, cc, data.gov.au and Schema in linked data and validates the value of these properties. |
| Average |  | 2 |  |
| Poor |  | 1 | The metadata does not contain an explicit license pointer. |

#### FM-Gen2-R1.1.2

#### 

|  |  |  |  |
| --- | --- | --- | --- |
| Levels | Color Representation | Score | Criteria |
| Exemplary |  | 3 | The linked data metadata contains an explicit pointer to the license. Tests: license attributes xhtml, dvia, dcterms, cc, data.gov.au and Schema in linked data and validates the value of these properties. |
| Average |  | 2 |  |
| Poor |  | 1 | The linked data metadata does not contain an explicit license pointer. |

Referências:

Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data <https://www.biorxiv.org/content/biorxiv/early/2017/03/20/117812.full.pdf>