EuPathDB: <https://eupathdb.org/eupathdb/>

Variable Control:

|  |  |
| --- | --- |
| **URI used** | <https://tritrypdb.org/tritrypdb/app/record/gene/LmjF.04.0130> |
| **Analysis date** | 29/04/2020 |
| **Acronyms** | |
| **RaCE -** Researcher Compliance Experiment | |
| **MaCE -** Machine Compliance Experiment | |

FINDABLE

|  |  |  |
| --- | --- | --- |
| **Principle** | F1 | |
| **Description** | (Meta)data is assigned persistent and globally unique identifiers. | |
| **RaCE** | **/ \* Globally Unique? \* /**  They are standardized, but no standardization documentation was found. The use of URI makes Globally Unique Identifiers .  **/ \* Persistent? \* /**  According to Re3Data, identifiers are not persistent.  https://www.re3data.org/repository/r3d100011557  The identifiers were not found on identifiers.org:  https://registry.identifiers.org/registry/uniprot  **/\*Comments\*/**  Local identifiers are standardized and documented. | |
| **Result of RaCE (by color)** |  | |
| **Recommendations** | There are no improvements. | |
| **MaCE** | **METRIC DESCRIPTION** | **ANALYZE** |
| **Unique Identifier**  Test whether the metadata resource has a unique identifier | It has an Uniform Resource Identifier type identifier . |
| **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 ). | The metadata GUID does not conform to any known permanent URL system. |
| **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%2 0schema). For URLs that do not follow a scheme in FAIRSharing , we tested known URL persistence schemes ( purl , oclc , fdlp , purlz , w3id, ark ). | It is not possible to find the data identifier in the metadata using any property / predicate ( common) reserved for this purpose. |
| **Result of MaCE (by color)** | **Unique Identifier** |  |
| **Identifier Persistence** |  |
| **Data Identifier Persistence** |  |
| **Recommendations** | **Unique Identifier** | There are no improvements. |
| **Identifier Persistence** | It is necessary to adopt an identifier persistence repository, such as those mentioned in the metric. |
| **Data Identifier Persistence** | It is necessary to adopt an identifier persistence repository, such as those mentioned in the metric. |

|  |  |  |
| --- | --- | --- |
| **Principle** | F2 | |
| **Description** | Data is described with rich metadata. | |
| **RaCE** | **/ \* Is there metadata standardization ? \* /**  Yes, the data are standardized.  **/ \* Is the metadata complete? \* /**  No, even with standardization all the metadata gaps appear, however they are empty without the metadata data.  **/ \* Comments \* /**  There is n't . | |
| **Result of RaCE (by color)** |  | |
| **Recommendations** | At first, there is a need to include only existing metadata, not allowing empty fields to be returned.  Another issue is related to better metadata documentation. | |
| **MaCE** | **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 . | No structured metadata found (biblio teca extruct in python and https://demos.algorithmia.com/web-page-inspector was used ) |
| **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, even when terms are spaced by name. Versions future of this test may be more , flexible s . | Extruct was unable to extract any kind of structured metadata. |
| **Result of MaCE (by color)** | **Structured Metadata** |  |
| **Grounded Metadata** |  |
| **Recommendations** | **Structured Metadata** | It is necessary to use structured data languages ​​to assist in the use of the data. |
| **Grounded Metadata** | It is necessary to use structured data languages ​​to assist in the use of the data. |

|  |  |  |
| --- | --- | --- |
| **Principle** | F3 | |
| **Description** | Metadata must clearly and explicitly include the data identifier described. | |
| **RaCE** | **/ \* Is it possible to identify the data in the metadata? How is it done? \* /**  It is possible, the data is on the same page as the metadata, however it is not possible to use scrapping tools and there is no connection between the two using the same metadata.  **/ \* Are there different identifiers for metadata and data? \* /**    There is not, it uses the same identifiers.    **/\*Comments\*/**    Reference other IDs external to the repository. | |
| **Result of RaCE (by color)** |  | |
| **Recommendations** | It is necessary to explain the identifiers to the metadata and data so that they can be connected, even in a simple way.  In order to further improve it, it would be essential to adopt means so that data extraction can be carried out via automation. | |
| **MaCE** | **Data Identifier Explicitly in Metadata**  Metric for testing 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 codeRepo sitory scheme is used for software versions . | No structured metadata found (the extruct library in python and https://demos.algorithmia.com/web-page-inspector was used ) |
| **Metadata Identifier Explicitly in Metadata**  Metric to test whether the metadata contains the unique identifier for the metadata itself. This is done using a variety of scraping tools , including resolving DOI metadata, using the Python tool ' extruct ' and other ... | Extruct was unable to extract any kind of structured metadata. |
| **Result of MaCE (by color)** | **Data Identifier Explicitly in Metadata** |  |
| **Metadata Identifier Explicitly in Metadata** |  |
| **Recommendations** | **Data Identifier Explicitly in Metadata** | The use of structured data together with the identification of properties in these structured data would bring more value to the (meta)data stored and thus meet this metric. |
| **Metadata Identifier Explicitly in Metadata** | The use of structured data together with the identification of properties in these structured data would bring more value to the (meta)data stored and thus meet this metric. |

|  |  |  |
| --- | --- | --- |
| **Principle** | F4 | |
| **Description** | (Meta)data is recorded or indexed in searchable resources. | |
| **RaCE** | **/ \* The URI or local ID returns the record in engines of search? \* /**  Yes, it is returned . | |
| **Result of RaCE (by color)** |  | |
| **Recommendations** | There are no improvements. | |
| **MaCE** | **Searchable in Major Search Engine**  It tests whether a machine is able to discover the resource by searching, using Google. | The searcher used the identifier and returned the record of (meta)data |
| **Result of MaCE (by color)** | **Searchable in Major Search Engine** |  |
| **Recommendations** | **Searchable in Major Search Engine** | There are no improvements. |

ACCESSIBLE

|  |  |
| --- | --- |
| **Principle** | A1 |
| **Description** | (Meta)data is retrieved by its identifier using standardized communication protocols. |
| **RaCE** | **/ \* Do database use standardized protocols? \* /**    Yes, it uses standardized protocols for data access.    **/ \* Do database has proprietary software for data access? \* /**    It does not have any proprietary software.    **/ \* Comments \* /**    There isn't . |
| **Result of RaCE (by color)** |  |
| **Recommendations** | There are no improvements. |
| **MaCE** | There is no FAIR METRICS GEN2 for this principle. |
| **Result of MaCE (by color)** | There is no FAIR METRICS GEN2 for this principle. |
| **Recommendations** | There is no FAIR METRICS GEN2 for this principle. |

|  |  |  |
| --- | --- | --- |
| **Principle** | A1.1 | |
| **Description** | Protocol is open, free and universally implemented. | |
| **RaCE** | **/ \* Does it have open, free and universally implemented access? \* /**    Yes, no registration or proprietary software is required to access (meta)data.  **/ \* Comment s \* /**    There may be inaccurate (meta)data due to the possibility of submitting (meta)data from ongoing research. | |
| **Result of RaCE (by color)** |  | |
| **Recommendations** | There are no improvements. | |
| **MaCE** | **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 . The recognition of other identifiers will be added at the request of the community. | No keys are passed (using the extract and analysis via JSON). |
| **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 . The recognition of other identifiers will be added at the request of the community. | No keys are passed (using the extract and analysis via JSON). |
| **Result of MaCE (by color)** | **Uses Open Free Protocol for data Retrieval** |  |
| **Uses Open Free Protocol for Metadata Retrieval** |  |
| **Recommendations** | **Uses Open Free Protocol for data Retrieval** | The use of URIs and CURIEs that provide more information on the (meta)data is essential in this matter because it raises the level of (meta)data present in the data. |
| **Uses Open Free Protocol for Metadata Retrieval** | The use of URIs and CURIEs that provide more information on the (meta)data is essential in this matter because it raises the level of (meta)data present in the data. |

|  |  |  |
| --- | --- | --- |
| **Principle** | A1.2 | |
| **Description** | The protocol allows for authentication and authorization when necessary. | |
| **RaCE** | **/ \* Does the protocol allow authentication and authorization? \* /**    There is no such restriction. | |
| **Result of RaCE (by color)** |  | |
| **Recommendations** | There are no improvements. | |
| **MaCE** | **Data Authentication and Authorization**  Test a discovered data GUID for the ability to implement authentication and authorization in resolution protocol. Currently passes InChI keys , DOIs , identifiers and URLs . He also searches for metadata by Dubl in Core's ' accessRights ' property , which can point to a document that describes the data access process. The recognition of other identifiers will be added at the request of the community. | The GUID returned from the metadata is a URI, known to allow authentication / authorization . |
| **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 . Recognition from other i dentifiers will be added at the request of the community. | The GUID returned from the metadata is a URI, known to allow authentication / authorization. |
| **Result of MaCE (by color)** | **Data Authentication and Authorization** |  |
| **Metadata Authentication and Authorization** |  |
| **Recommendations** | **Data Authentication and Authorization** | There are no improvements. |
| **Metadata Authentication and Authorization** | There are no improvements. |

|  |  |  |
| --- | --- | --- |
| **Principle** | A2 | |
| **Description** | Metadata must be accessible even when data is no longer available. | |
| **RaCE** | **/ \* Is there a data versioning policy? \* /**  There is the storage of the IDs of the records, they are allocated in the *previous ID* , when trying to access these IDs through a URI, they return a nonexistent page.  **/ \* (Meta)data can be erased? \* /**  Yes, they can be deleted.  **/ \* (Meta)data can be updated? \* /**  Yes, they can be updated.  **/ \* Is it possible to access metadata of data that no longer exists? \* /**  It is not possible, return Page Not Found  **/ \* Is there a persistence policy? \* /**  No persistence policy was found.  **/ \* Comments \* /**  There are no comments. | |
| **Result of RaCE (by color)** |  | |
| **Recommendations** | Even if there is an indication of previous IDs, implying that there are updates, the old URIs give an error when accessing, a possible solution to meet this metric is instead of returning Page Not Found, redirecting to the data with newer ID. | |
| **MaCE** | **Metadata Persistence**  Metric for testing whether 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. | It was not possible to find any persistence policy using any approach |
| **Result of MaCE (by color)** | **Metadata Persistence** |  |
| **Recommendations** | **Metadata Persistence** | In order to comply with this metric, it is necessary to indicate a persistence policy in the registration source code. |

INTEROPERABLE

|  |  |  |
| --- | --- | --- |
| **Principle** | I1 | |
| **Description** | (Meta)data use formal, accessible, shared and widely applicable language for knowledge representation. | |
| **RaCE** | **/ \* Is there a use of languages ​​to represent knowledge? \* /**    Languages ​​are not used to represent knowledge.    **/ \* Taxonomies, ontologies, controlled vocabularies are referenced or found in the repository? (Must meet F 1) \* /**    Controlled vocabularies were found there quote from ontologies, but the test record cites the ontology but they do not exist.  **/\*Comments\*/**    There is a submission protocol ( [https://eupathdb.org/EuPathDB\_datasubm\_SOP.pdf](https://translate.google.com/translate?hl=pt-BR&prev=_t&sl=pt&tl=en&u=https://eupathdb.org/EuPathDB_datasubm_SOP.pdf) ) but, until now, no structured means for submitting metadata have been found. | |
| **Result of RaCE (by color)** |  | |
| **Recommendations** | Use some type of representation language in the records so that they can be easily analyzed, develop documents to assist in understanding the metadata for users. | |
| **MaCE** | **Metadata Knowledge Representation Language (WEAK)**  Maturity indicator to test 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. | There are no knowledge representation languages ​​in the register used in the research. But there is a section containing “some” regions mapped with knowledge representation languages. |
| **Metadata Knowledge Representation Language (STRONG)**  Maturity indicator to test whether metadata uses a formal language widely applicable to knowledge representation. This particular t this 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 ontology s. Any form of RDF will pass this test (including RDF that is automatically extracted by third-party analyzers, such as Apache Tika ). | It does not meet this metric. |
| **Result of MaCE (by color)** | **Metadata Knowledge Representation Language (WEAK)** |  |
| **Metadata Knowledge Representation Language (STRONG)** |  |
| **Recommendations** | **Metadata Knowledge Representation Language (WEAK)** | The use of data representation languages ​​is essential to increase the reuse of data and assist in the formation of linked data. |
| **Metadata Knowledge Representation Language (STRONG)** | The use of data representation languages ​​is essential to increase the reuse of data and assist in the formation of linked data. |

|  |  |  |
| --- | --- | --- |
| **Principle** | I2 | |
| **Description** | (Meta)data using the following vocabulary FAIR Principles. | |
| **RaCE** | **/ \* Vocabularies following the FAIR Principles? \* /**    They don't follow.  **/\*Comments\*/**    There isn’t. | |
| **Result of RaCE (by color)** |  | |
| **Recommendations** | One of the improvements that would help the FAIRification process is to use FAIR Principles in the vocabularies, as they support the entire repository, thus facilitating future phases of FAIRification .    Using FAIR DATA POINT can be a way of tending this Principle. | |
| **MaCE** | **Metadata Uses FAIR Vocabularies (WEAK)**  Maturity indicator to test whether linked data metadata uses resolved terms. This only tests if they resolve, and FAIR data does not resolve, so it is a rather weak test. | Does not meet this metric using Extruc t , https://demos.algorithmia.com/web-page-inspector |
| **Metadata Uses FAIR Vocabularies (STRONG)**  Maturity indicator to test whether linked data metadata uses resolved linked data (FAIR) terms. | No resolved predicates were found for linked data. |
| **Result of MaCE (by color)** | **Metadata Uses FAIR Vocabularies (WEAK)** |  |
| **Metadata Uses FAIR Vocabularies (STRONG)** |  |
| **Recommendations** | **Metadata Uses FAIR Vocabularies (WEAK)** | The use of representation languages ​​would assist in the identification of resolvable links and thus meet this metric. |
| **Metadata Uses FAIR Vocabularies (STRONG)** | For linked data to be used, for this to occur, data must be written in knowledge representation languages.  Use of FAIR DATA POINTS would assist in meeting this metric. |

|  |  |  |
| --- | --- | --- |
| **Principle** | I3 | |
| **Description** | (Meta)data includes qualified references for other (meta)data. | |
| **RaCE** | **/ \* Is there use of structured data languages ? \* /**    Do not use.    **/ \* Is there qualification among (meta)data entities ? \* /**    There is no establishment of qualifications.    **/ \* Comments \* /**    There is n't . | |
| **Result of RaCE (by color)** |  | |
| **Recommendations** | The use of structured data languages ​​would achieve this Principle considering that these types of languages ​​provide a means for characterizing qualification among (meta)data. | |
| **MaCE** | **Metadata Contains Qualified Outward References**  Maturity indicator to test whether metadata is externally linked to third-party resources. It only tests metadata that can be represented as linked data. | It is not possible to establish this connection with linked data |
| **Result of MaCE (by color)** | **Metadata Contains Qualified Outward References** |  |
| **Recommendations** | **Metadata Contains Qualified Outward References** | The use of representation languages ​​would assist in achieving this metric, since from its use it provides a means for implementing linked data. |

REUSABLE

|  |  |
| --- | --- |
| **Principle** | R1 |
| **Description** | (Meta)data are richly described with the plurality of precise and relevant attributes. |
| **RaCE** | **/ \* Is there a standardization of (meta)data? \* /**    It exists, but it is not complete.    **/ \* Is there reference to metadata in other repositories? \* /**    Yes, there is this type of reference.    **/ \* Comments \* /**    There is n't . |
| **Result of RaCE (by color)** |  |
| **Recommendations** | Even though it contains a lot of (meta)data, or indication of it, there are a lot of gaps to be filled from the example record. |
| **MaCE** | There is no FAIR METRICS GEN2 for this principle. |
| **Result of MaCE (by color)** | There is no FAIR METRICS GEN2 for this principle. |
| **Recommendations** | There is no FAIR METRICS GEN2 for this principle. |

|  |  |  |
| --- | --- | --- |
| **Principle** | R1.1 | |
| **Description** | (Meta)data is published with clear and accessible data usage licenses. | |
| **RaCE** | **/ \* Usage licenses are found in the registry? \* /**    Do not.    **/ \* Are usage licenses found in the repository? \* /**    Yes.    **/ \* Link to user licenses \* /**    [https://www.re3data.org/repository/r3d100011557](https://translate.google.com/translate?hl=pt-BR&prev=_t&sl=pt&tl=en&u=https://www.re3data.org/repository/r3d100011557)  However, Re3Data cites non-existent licenses. | |
| **Result of RaCE (by color)** |  | |
| **Recommendations** | The use licenses are found in the repository, however they are not referenced in the records. | |
| **MaCE** | **Metadata Includes License ( Weak )**  Maturity indicator to test whether metadata contains an explicit license pointer. This test 'weak' will use a regular expression that differs from case sensitive and will sweep the style metadata key / value, as well as metadad the linked data. Tests: xhtml , dvia , dcterms , cc , data.gov.au and Schema license predicates on linked data and validates the value of these properties. | No explicit pointers to Machine-Readable referring to the use license were found. |
| **Metadata Includes License (Strong)**  Maturity indicator to test whether the linked data metadata contains an explicit pointer to the license. Tests: xhtml , dvia , dcterms , cc , data.gov.au and Schema license predicates on linked data and validates the value of these properties. | No explicit pointers to Machine-Readable referring to the use license were found. |
| **Result of MaCE (by color)** | **Metadata Includes License (Weak)** |  |
| **Metadata Includes License (Strong)** |  |
| **Recommendations** | **Metadata Includes License (Weak)** | As there are no knowledge representation languages, it is very complicated to find any kind of license predicate without this representation, so it is essential to use it . |
| **Metadata Includes License (Strong)** | As there are no knowledge representation languages, it is very complicated to find any kind of license predicate without this representation, so it is essential to use it . |

|  |  |
| --- | --- |
| **Principle** | R1.2 |
| **Description** | (Meta)data are associated with detailed provenance. |
| **RaCE** | **/ \* Provenance (meta)data is identified? \* /**    Yes, the source metadata is identified.    **/ \* There are documents to describe the provenance \* /**    There are no detailed documents of provenance. |
| **Result of RaCE (by color)** |  |
| **Recommendations** | Even if the (meta)data is associated with detailed provenance, there is no connection, nor documents that refer to the conception of a provenance. |
| **MaCE** | There is no FAIR METRICS GEN2 for this principle. |
| **Result of MaCE (by color)** | There is no FAIR METRICS GEN2 for this principle. |
| **Recommendations** | There is no FAIR METRICS GEN2 for this principle. |

|  |  |
| --- | --- |
| **Principle** | R1.3 |
| **Description** | (Meta)data meet community standards relevant to the domain. |
| **RaCE** | It is not possible to analyze this Principle. |
| **Result of RaCE (by color)** | It is not possible to analyze this Principle. |
| **Recommendations** | It is not possible to analyze this Principle. |
| **MaCE** | There is no FAIR METRICS GEN2 for this principle. |
| **Result of MaCE (by color)** | There is no FAIR METRICS GEN2 for this principle. |
| **Recommendations** | There is no FAIR METRICS GEN2 for this principle. |