Key Recommendations to improve FAIR Principles adaptation (References listed provide specific details and more extensive discussion on these topics)

## Findable

#### F1

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - Make your (meta)data persist over time. Provide persistent and globally unique identifiers that are standardized, can be done through the creation of own models (obeying persistence rules) or the use of highly consolidated models and services using identifier persistence schemes.  - Use existing standards URI-style identifier formats (compressed or normal). | <https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/>  <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2001414>  <https://www.sciencedirect.com/science/article/pii/S2666389920300040>  <https://www.sciencedirect.com/science/article/pii/S1567422311000251>  <https://www.w3.org/Addressing/URL/uri-spec.html>  <https://www.ietf.org/rfc/rfc3986.txt>  <https://whatis.techtarget.com/definition/URI-Uniform-Resource-Identifier>  <https://link.springer.com/referenceworkentry/10.1007%2F978-1-4419-9863-7_1572>  <https://www.w3.org/TR/2010/NOTE-curie-20101216/>  <https://www.crossref.org/blog/curies-a-cure-for-uris/>  <https://www.crossref.org/blog/curies-a-cure-for-uris/>  <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2001414>  <https://www.sciencedirect.com/science/article/pii/S1567422311000251>  <https://www.sciencedirect.com/science/article/abs/pii/S1574954116300140>  <https://www.sciencedirect.com/science/article/pii/S1368837519303021>  <https://www.sciencedirect.com/science/article/abs/pii/S0167739X18314638> |

#### F2

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - The metadata must be as complete as possible. Control on the insertion of metadata must be high mainly so that there’s no lost metadata when the data holder inserts into the repository, a metadata upload policy is essential, so that there is no loss, inconsistency, inconsistency of metadata.  - Creating a grounded metadata standardization is considered essential in this part, using existing ontologies, vocabularies, thesaurus or creating new ones helps in increasing the richness of metadata.  - Information about data, such as its description, description of columns, meaning of abbreviations, units of measure, figures, analysis of the data itself, how the data was collected and / or suggestions on how to reuse the data are examples of essential metadata. Another recommendation factor is the creation of documentation to help users with data to know about the requirements for submitting (meta)data and, in addition, to make it as intuitive as possible using processes already grounded in the area of ​​Human Computer Interaction.  - Use structured data that is compatible for running applications focused on data scraping, languages like RDFa, embedded json, json-ld or structured content-negotiated metadata, such as RDF Turtle, are examples of languages that support data scraping.  - Use a structured metadata. There are numerous means of standardizing metadata structure, where they are located in the source code of the page, or by means of an extension and designated for structured data.  - Make the use of grounded metadata, which refers to the use of data standardization through languages recognized by computers for data disposal, etc. However, in this case, the purpose of further increasing the automation of these applications is used to use named nodes and use blank nodes, that is, use identifiers (persistent) to use characters (dbpedia is a good example where to find persistent identifiers for names) | <https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/>  <https://theodi.org/knowledge-opinion/guides/>  <https://isa-tools.org/>  <https://www.dtls.nl/fair-data/find-fair-data-tools/>  <https://books.google.com.br/books?id=oY1lAgAAQBAJ&pg=PT43&lpg=PT43&dq=grounded+rdf&source=bl&ots=6TMIKsE9_0&sig=ACfU3U1auvNc_TNB1jh78KvbxH9QYkyCWQ&hl=pt-BR&sa=X&ved=2ahUKEwilyLiAr5nqAhVIDrkGHTGzBY0Q6AEwAHoECAYQAQ#v=onepage&q=grounded%20rdf&f=false> |

#### F3

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - It is essential to explain the identifiers of the datasets (data) to the metadata so that they can be connected, even in a simple way. Use of persistent and globally unique identifiers (F1) helps to achieve this principle. In order to further improve it, it would be essential to adopt a means to extract data via automation.  - It is necessary to have some data structure that is recognized by machine (eg, RDFa, embedded json, json-ld or structured metadata negotiated by content, such as RDF Turtle), in addition, to have specifications / properties / tags in order to locate the identifier of the data (s) easily in the metadata.  - The more you use resolvable metadata (using URIs) the better it will be for automatic data collection applications to understand about the (meta) data present in that repository. Again, the use of structured data standardization that is recognized by machines will be better reuse. |  |

#### F4

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - One of the possible improvements is the standardization of local identifiers that assist in search engines, considering that from their crawlers they would identify Local IDs as well as URIs.  - Free access to crawlers, standardization and ID documentation is essential to meet this metric. |  |

## Accessible

#### A1

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - Provide means so that (meta)data can be accessed without the need for proprietary software and protocols. Use widely known protocols (e.g. http, https, ftp ...) and not use software and standards for non-proprietary files (e.g. use csv instead of excel files). In addition, it is seen the need to document and demonstrate to the user what the process is for downloading data, as well as making the process as intuitive and accessible as possible (Human Computer Interaction). |  |

#### A1.1

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - In the case of accessing data that do not have access restrictions, it is essential that they are accessed through non-proprietary protocols, do not offer any type of restriction for accessing the data, provide means for understanding how the data can be accessed without restriction (eg documents, tutorials).  - Use of URIs (e.g. InChI keys, DOIs, identifiers and URLs) within the metadata that return the data itself and information about the metadata without the need for authentication and authorization and the use of proprietary protocols or software. |  |

#### A1.2

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - If there is a need to restrict the (meta)data (e.g. sensitive data related to health) it is necessary to establish means for there to be restriction through user authentication and authorization through the data holder or repository. It is extremely important that the entire process for accessing the data is documented for the understanding of the user who needs to use the data in question.  - Provide means to make it possible to carry out authentication and authorization automatically for data access or to identify that users must go through previous authentication and / or authorization steps to access data (eg use tags in languages ​​recognizable by machines that identify this need).  - One must take into account the need to provide some type of free access metadata, in order to assist the user in having knowledge of the data that he will have access to, in order for the user to be sure that it is the data he needs, thus being able to optimize the functions both on the user side and on the data holder and repository side. The same recommendations applied to FMGEN2 - Data Authentication and Authorization apply to this recommendation. |  |

#### A2

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - Creation of versioning policies, use of globally unique and persistent identifiers with the storage of old versions of metadata. Metadata in text type, link to the most up-to-date data, metadata about the data holder, link to versioning repository containing old metadata, are examples of improvement.  - Resolution of possible URIs prior to (meta)data so that the user is not led to errors in access (resolution to return to the newest URI). Pay attention to this principle mainly due to the F4 where the old URI may be indexed with a higher value than the new URI.  - Use of persistence and versioning policies are essential to not generate problems to the (meta)data (e.g. inconsistency), when using policies at this level it is necessary that they are seen by machines, in order to assist in understanding where to find the newer (meta)data if old versions are accessed, or that it is no longer available. |  |

## Interoperable

#### I1

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - Provide means for the (meta)data to be as detailed as possible regarding the use of knowledge representation languages (eg RDF, OWL, DAML-OIL, JSON LD, are highly used examples) and the use of thesaurus, ontologies, controlled vocabularies, where they meet the requirement F1. Compliance with these recommendations helps to increase interoperability between different (meta)data. |  |

#### I2

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - This recommendation is fulfilled from the adaptation or development of vocabularies used by the repositories to the FAIR principles, the standardization can take place through the adoption of the FAIRification Process and adoption of the FAIR Data Point. |  |

#### I3

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - Qualified references do not say about the quality of the data, which may result in different understandings, but in the provision of standardized identifications of their references. Qualified references are achieved through the use of knowledge representation languages and the use of globally unique and persistent identifiers, thus being based on standards to obtain grounded (meta)data and consequently qualifications between the possible relationships between (meta)data. Use of external URIs in knowledge representation languages with qualification tags. |  |

## Reusable

#### R1

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - Completeness of (meta) data facilitates its reuse. The more diversification of (meta)data the better for its reuse. (Meta)data involving the capture, analysis, refinement, instructions for use, processing, software and hardware used are essential for understanding the stored data. It is essential to have discussions among stakeholders about which (meta)data should be essential for the data to be published / stored. |  |

#### R1.1

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - Use and abuse in well-founded use licenses (e.g. Creative Commons Licenses) to determine the responsibilities of (meta)data reuse, especially for sensitive/restricted data. If the (meta)data are freely accessible, they must be based on free use licenses.  - It is necessary to include in the records means to be able to identify the use licenses. Use of license xhtml, dvia, dcterms, cc, data.gov.au and Schema in linked data are examples to meet this metric. |  |

#### R1.2

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - Use of provenance as well as retrospective with a focus on documenting the life cycle of digital objects, use of proven provenance models to extract what data is essential for the reuser, there are provenance systems that facilitate compliance with this recommendation. |  |

#### R1.3

|  |  |  |
| --- | --- | --- |
| **Field** | **Recommendation** | **References** |
| **FAIR Principles** | - Acceptance by the standardization of communities relevant to the theme of the (meta) data ends up adding several benefits to the (meta) data itself and to the repository. It is necessary to research what existing standards exist for the theme, and possibly to implement data control to be inserted by the holders of (meta) data. |  |

#### 