**Wheat Data Semantic Interoperability. Overview on Metadata and Ontologies used in the Wheat Research Data Community**

Draft 1.0

**Carme Reverté Reverté and Imma Subirats [[1]](#footnote-1)**

# Wheat Data Interoperability WG (Research Data Alliance)

Contents

[Wheat Data Interoperability WG (Research Data Alliance) 1](#_Toc399518257)

[1. Introduction 2](#_Toc399518258)

[2. Data analysis and semantic interoperability 3](#_Toc399518259)

[2.1. Data structures 3](#_Toc399518260)

[2.1.1. Analysis 3](#_Toc399518261)

[2.1.2. Who are the users 5](#_Toc399518262)

[2.1.3 Recommendations 8](#_Toc399518263)

[3. Ontologies 9](#_Toc399518264)

[3.1. Analysis 9](#_Toc399518265)

[3.2. Results 12](#_Toc399518266)

[4. Recommendations 18](#_Toc399518267)

[4.1. How could the semantic interoperability be improved? 18](#_Toc399518268)

[5. Bibliography 19](#_Toc399518269)

[Annexes 20](#_Toc399518270)

[Annex 1. Mapping of three formats: structure and semantic differences 20](#_Toc399518271)

[Annex 2. Ontologies analyzed 22](#_Toc399518272)

# 1. Introduction

The use of common metadata, ontologies and derived standards are an essential part of the accessibility and discovery of datasets in international scientific communities. In the context of wheat research, a wide range of datasets is currently produced, with a consequent need to define features and scenarios to assess the level of interoperability.

This report is based on the “Analysis of survey: data standards in the wheat research community wheat data interoperability WG”[[2]](#footnote-2) (RDA, 2014). Semantics seek to enhance the interoperability of data through the use of metadata and ontologies as much standardized as possible. This study analyzes the semantics currently in use in the context of wheat research data like type of data, data descriptors used by scientists or in information systems to represent data, and mechanisms or systems for data interchange.

It is focused on the two main pillars of the semantic interoperability: data structures and controlled vocabularies (including entities such as institutions, geographic names, etc)[[3]](#footnote-3). Following this methodology, we have analyzed the metadata and ontologies used in the wheat research data community, present in the survey, describing its semantic interoperability. The main objective is to assess the level of compatibility and harmonization of data.

# 2. Data analysis and semantic interoperability

Two different types of information have been analyzed from the “Analysis of survey: data standards in the wheat research community wheat data interoperability WG” (RDA, 2014): data structures and controlled vocabularies.

## 2.1. Data structures

In the case of **data structures,** a wide number of metadata formats have been identified and grouped by type of data:

* SNPs,
* Genomic
* Annotations,
* Phenotypes,
* Genetic Maps,
* Physical Maps,
* Germplasm, and
* Gene Expression.

For each group of data, the survey identified specific formats as a core models[[4]](#footnote-4). These core models are taken as samples for this semantic interoperability analysis. We haven’t included a complete analysis of these data exchange formats; we do not have access examples of records. Additionally to have a full comprehension of the use of these formats is required some more analysis together with final users.

### 2.1.1. Analysis

The formats analyzed are the following:

|  |  |  |  |
| --- | --- | --- | --- |
| **Type of data**  | **Data structures or metadata formats used** | **Comments** | **Representation formats** |
| SNPs | BAM/SAM; VCF, VARSCAN | Information extracted from their websites[[5]](#footnote-5) | ASCII |
| Genomic annotations | GenBank format Flat | Information extracted from their websites[[6]](#footnote-6).  | TXT |
| Phenotypes | Isa-tab | Information extracted from a record in Isa-tab web page.[[7]](#footnote-7) | TXT and XML |
| Germplasm  | MCPD | Biodiversity Information standards website.[[8]](#footnote-8) . See also an article published By FAO in 2012[[9]](#footnote-9) and the Developing crop descriptors lists[[10]](#footnote-10).  | XML  |

Table 1. Data structures and formats analyzed

Two different elements were analyzed related to the semantic interoperability study: use of representation formats for wheat metadata and labels or descriptors used by each metadata format.

Essentially, the analysis shows that there is lack of information about the use of markup languages for each type of data. Most of them are represented by txt. Although, the group of SNPs (the most common language) uses ASCII, that could be converted to RDF. Furthermore, XML is used in Germplasm and Phenotypes core formats.

Comparing the metadata labels or information descriptors, we concluded that there is lack of common terminology among the different metadata formats analyzed. Several kinds of data require also different description of data, but all metadata have some common labels related to the identification of an object. Then, in each metadata format there is a group of labels to identify the object described (location information are considered as a part of this group) and the rest of labels are used to describe the object or specific data (genotypes, etc.). See Annex 1.

Differences and similarities among formats:

|  |  |
| --- | --- |
| **Differences** | **Similarities**  |
| * Different terminologies: different ways to label the same concept
* Different properties and definitions
* Different structures: e.g. GenBank format includes categories and subcategories of labels, the rest of metadata includes a unique level of descriptors
* Use of different formats for data representation: ASCII, txt and xml
* Differences in content and coverage
 | * Good level of interoperability with plane formats: txt and ASCII
* Comprehensive formats or human readable formats
* Use of similar labels or properties for data identification (e.g. reference name and position of the object)
* Similar interests (related topics)
 |

Table 2. Differences and similarities among formats

### 2.1.2. Who are the users

When the objective is to provide good practices on metadata, it is important to know what and who is using it, and in what context. In this section we will not go through the data already present in the survey again, especially on who participated in the survey, and their profiles. However, we would like to show in more detail the differences on type of usage for each data format.

The distinction between consumption and production of data is relevant to establish recommendations on good practices for exchange of data and metadata standards. First, we do not want to change specific synergies that might work fine in the community, but to accommodate the recommendations as much as possible to them. Second, it is easier to recommend metadata standards for the exchange of data, rather than for creating and managing data, therefore it is relevant to know what metadata formats are being used for the consumption of data.

The tables below present the usage of the data formats by 1.) user/consumer, 2.) producer and 3.) both user/consumer and producer, all listed by data type. In general, there are strong similarities on what metadata good practices are used for exchange of data and data management. However, there are also exceptions, as in the case of data type germplasm.

Figure 1. Usage of data formats for data type SNPs

Figure 2. Usage of data formats for data type Genomic annotations

Figure 3. Usage of data formats for data type phenotypes

Figure 4. Usage of data formats for data type genetic maps

Figure 5. Usage of data formats for data type physical maps

Figure 6. Usage of data formats for data type germplasm

Figure 7. Usage of data formats for data type gene expression

### 2.1.3 Recommendations

Harmonization is a priority. It is necessary to develop common set of terminologies, basic properties and structure and processes for metadata description and management. As an example of the importance of doing so, it is important to highlight that not all formats present in the survey have the unique identifier property, an essential property in the context of research data.

Some recommendations for the harmonization of metadata:

* **Terminology**. It is recommended to define a common terminology for the specification of, at least, a general metadata use (e.g. Dublin Core versus Darwin Core) and its definition. A shared vocabulary helps for harmonizing efforts.
* **Properties**. Some similarities are found in labels related for object identification. So, the similarities need to be extracted and the concepts generalized and used in a common way as part of metadata standardization. Examples of common properties.
	+ Requiring the use of unique identifiers for each metadata element (e.g. label, identifier, field name, etc.)
	+ Defining each metadata element using common semantics
	+ Establishing what elements are mandatory and if they are repeatable
	+ Structuring metadata elements: harmonize the use of hierarchical relations (GenBank and MCPD use hierarchical labels: classes and subclasses of elements)
	+ Defining the constraints of each element (e.g. MCPD has some restrictions in certain element as ACQDATE or Acquisition date, that is represented as YYYYMMDD)
* **Structures**. Mapping is recommended as the different metadata standards are organized differently. Although, crosswalk processes are complicated by the fact that each format use a different organization, the homogenization process helps to develop specific and standardized requirements for content-specific elements of different metadata formats. It is recommended a deep analysis or intellectual tasks of potential mappings or crosswalk processes among metadata formats including such semantic as system analysis.

## 3. Ontologies

We have analyzed 15 ontologies, the most used in the survey[[11]](#footnote-11). The analysis is based on those used by this community and the accessibility of their data itself (its representation and comprehensibility for humans and machines)[[12]](#footnote-12).

We have selected the most important indicators to analyze the interoperability level of each ontology: **data** **description** (title, creator, publisher and URL, type of data content and language), **use of standards** (representation standards, metadata standards, etc.) **data access for information exchange** (protocols for data sharing, technical accessibility and license) and an additional attribute was also included, the **interlinked level** of each controlled vocabulary as an added value to data quality[[13]](#footnote-13). The analysis also includes statistical data.

## 3.1. Analysis

Only 53% of the analyzed ontologies could be describe using the indicators described above (Figure 1) .



Figure 8.Ontologies with a minimum of data quality (See Annex 2 for more information about the ontologies)

Quality and interoperability are essential in this analysis; therefore we have developed a metric for each indicator evaluation using data interoperability parameters: **identification of data** (data description), **use of formats and standards** (data representation), **accessibility** and **interlinking**.

Each group of indicators has a different level of relevance in terms of semantic interoperability[[14]](#footnote-14). In this case, we proposed to give more value to the most used standards and relevance to accessibility parameters as the most important values to achieve the semantic interoperability. However, a further analysis is required to build a data quality framework for wheat ontologies and related fields.

Quality characteristics of the most used ontologies have been identified, and based on that we provide recommendations on good practices and recommendations for the data managers. However, this study only gives a general feedback. For a complete analysis it is necessary to analyze each of them in more detail.

Table 1 lists the quality criterion and suggested method for wheat ontologies interoperability assessment:

|  |  |  |
| --- | --- | --- |
| **Data quality indicator** | **Attributes[[15]](#footnote-15)** | **Scores[[16]](#footnote-16)** |
| **Data descriptors.** It is evaluated the absence or presence of an indicator. Absence is an empty value with 0 point, presence is a fill value with 1 point  | title  | 1 |
| creator | 1 |
| publisher | 1 |
| Url | 1 |
| type of data | 1 |
| content | 1 |
| language | 1 |
| **Total** | **7** |
| **Formats and Standards.** This indicator requires a more qualitative analysis than other indicators and the values are given taken into account the use of common standards.  | Representation standard formats (e.g. W3C standards)[[17]](#footnote-17) | OBO | 2 |
| RDF | 2 |
| RDF-XML | 2 |
| OWL/SKOS | 2 |
| SOFA | 2 |
| Graph formats | 2 |
| **Total** | **12** |
| Other representation formats (related to the scientific area) | GAF | 1 |
| GPAD | 1 |
| GO conventions | 1 |
| Gene Annotation | 1 |
| Annotation Association files | 1 |
| Others… | 1 |
| **Total** | 6 |
| Metadata standards: use of specific metadata to data description (e.g. D.C., Darwin Core, etc. | Dublin Core (DC) | 2 |
| CG Core (based on DC) | 2 |
| CFV  | 2 |
| Others… | 2 |
| **Total** | 8 |
| **Accessibility[[18]](#footnote-18)** | Technical accessibility (as more open liked is the access point, more valuable is it). And each element could appear together  | Endpoint  | 10 | 10 |
| Web services | 7 |  |
| Repositories (the ontology is part of a repository) | 7 |  |
| Databases (the ontology is accessible through an online database) | 7 | 7 |
| FTP  | 7 |  |
| URL | 5 | 5 |
| Download | 1 | 1 |
| Others | … |  |
| **Total** | 23 |
| Protocols for data sharing (As more standardized more value) | SPARQL endpoint | 10 | 10 |
| OAI-PHM | 8 |  |
| DAS Server  | 6 |  |
| GP1.1 | 6 | 6 |
| Others | … |  |
| **Total** | 16 |
| License of use (as more open is the licence more value has the attribute)  | Open License standardized (e.g. Creative Commonce) | 10 | 10 |
| Open licence (owner licence) | 8 |  |
| Open licence with restrictions  | 5 |  |
| No licence | 0 |  |
| Others | … |  |
| **Total** | 10 |
| **Interlinking**  | Linked to others controlled vocabularies (KoSs or linked data technology) | 10 | 10 |
| Linked to other controlled vocabularies (without linked data technology) |  5 |  |
| Linked to other databases | 3 | 3 |
| Others | … |  |
| **Total** | 13 |

Table 3. Wheat Ontologies interoperability assessment proposal

### 3.2. Results

Six of the resources present in the survey have been discarded in this analysis due to the fact that they are not ontologies (identified with the following numbers 8, 9, 10, 13, 14 and 15, see Annex 2).

All ontologies are using the same ontological language for its representation and development, the OBO format. OBO format can be represented with the OWL description logic language, so, OBO is compatible with OWL format. Itis also a collaborative experiment involving developers of science-based ontologies who are establishing a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain. In OBO project you can find mappings among other related ontologies[[19]](#footnote-19). For example, in the wheat field there is Gene Ontology (GO), which includes 4 ontologies which are linked among them. The GO project has developed three structured, controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner.

Figure 9. Use of standards for ontologies representation

Most of the analyzed ontologies are managed and accessible in OBO and OWL-RDF/XML.

The use of the standard formats for ontological representation supports the interoperability between ontologies and among information systems. An example of this interoperability is in the interlinking level of the ontologies analyzed. The analysis of the interlinked level of these ontologies shows that it is very high level of interoperability. So, it is possible to find links among the same field of ontologies or among ontologies of related fields. Furthermore, most of these ontologies analyzed are used in several databases, as is exposed in the following figure:



Figure 10. Interlinking level among ontologies used in wheat field

Additionally, the main metadata used for content annotation is not using standards. Even though in the case of Gene they are standardized, no ISO standard is used. However, they are represented in standards formats[[20]](#footnote-20) like OBO flat file format developed by the [Gene Ontology Consortium](http://geneontology.org/).

Figure 11. Annotation data formats in wheat and related fields

Figure 12. Metadata standards used in ontologies for wheat or related fields

Despite the differences in the use of metadata, this comparative analysis demonstrates that these ontologies, as a first level, are interoperable in their formats and they could be integrated in the same system, or used in other systems as part of the knowledge enrichment.

Although it is needed a deeper analysis to conclude this assertion (such as coverage and structure or relationships study among ontologies), in a second level, we discovered that the accessibility and availability of these ontologies was very low. Most of them have not an access endpoint and the most common access procedure it is the downloading, constraining the automatically information sharing with other data systems. Additionally, there is lack of information about the licensing of these ontologies.

Figure 13. Ontologies accessibility

Figure 14. Use of protocols for data sharing

Figure 15. Use of licenses in wheat ontologies

# 4. Recommendations

As a conclusion, the metadata used in the context of the wheat research data requires harmonization and the use of interoperable standards such as RDF. With regard to ontologies, standardized and interoperable standards are being used for its representation, facilitating the linkage to other information resources.

### 4.1. How could the semantic interoperability be improved?

* Promote the use of interoperable standards for data representation (such as XML, XML-RDF, RDF, etc.). As a strongly recommendation is the use of RDF for data structures instead of txt or plain text.
* Harmonize the data structures. It is necessary to develop a generic framework to revise the existing metadata standards. Mappings and crosswalks are recommended among these several metadata formats.
* Use of standards to access data and information exchange: use of open exchange protocols (Endpoints, OAI-PMH), open licenses (creative commons licenses), ontology copyright licensing, etc.

# 5. Bibliography

Doerr, M. 2003. The CINDOC Conceptual Reference Model: An Ontological Approach to Semantic Interoperability data. *AI Magazine,* V24(3), p.75-92

Naumann, F., Rolker, C., (2000). Assessment methods for information quality criteria. *Proceedings of the International Conference on Information Quality (IQ)*. *(Cambridge, MA, 2000),* p. 148-162.

Patel, M., Koch, T., Doerr, M. and Tsinaraki, C. (2005). Semantic Interoperability in Digital Library Systems. Other. UKOLN,University of Bath.

St. Pierre, M., LaPlante, B. (1999). Issues in Crosswalking Content Metadata Standards. *Information Standards* *Quaterly*, 1999, V.11(1), p.2-5

Wang, R.Y.; Strong, D. M. (1996). Beyond accuracy: what data quality means to data consumers. *J. Manage. Inf. Syst.*, March, 1996, 12 (4), p. 5-33

# Annexes

## Annex 1. Mapping of three formats: structure and semantic differences

|  |  |  |
| --- | --- | --- |
| **VARSCAN Headers (E.G. GERMINE)** | **VCF Headers** | **SAM/BAM Headers** |
|

|  |  |
| --- | --- |
| **Field** | **Description** |
| **Chrom**IDENTIFICATION | Chromosome or reference name |
| **Position** | Position from pileup (1-based) |
| **Ref** | Reference base at this position |
| **Cons** | Consensus genotype or variant called |
| **Reads1** | Number of reads supporting reference |
| **Reads2** | Number of reads supporting variant |
| **VarFreq** | Allele frequency of variant by read count |
| **Strands1** | Number of strands on which reference observed (0-2) |
| **Strands2**CONTENT | Number of strands on which variant observed (0-2) |
| **Qual1** | Average base quality of reference-supporting bases |
| **Qual2** | Average base quality of variant-supporting bases |
| **Pvalue** | P-value from Fisher's exact test (0.98 means not calculated) |
| **MapQual1** | Average mapping quality of reference-supporting reads |
| **MapQual2** | Average mapping quality of variant-supporting reads |
| **Reads1Plus** | Number of reference-supporting reads in + orientation |
| **Reads1Minus** | Number of reference-supporting reads in - orientation |
| **Reads2Plus** | Number of variant-supporting reads in + orientation |
| **Reads2Minus** | Number of variant-supporting reads in - orientation |
| **VarAllele** | Most prevalent variant allele (even if site not not called variant) |

 |

|  |
| --- |
| **Field** |
| **Chrom**  |
| **POS position** |
| **REF reference base(s)** |
| **Cons** |
| **Reads1** |
| **VarFreq** |
| **Strands1** |
| **Strands2** |
| **Qual1** |
| **Qual2** |
| **Pvalue** |
| **MapQual1** |
| **MapQual2** |
| **Reads1Plus** |
| **Reads1Minus** |
| **Reads2Plus** |
| **Reads2Minus** |
| **VarAllele** |

 |

|  |  |
| --- | --- |
| **Field** | **Description** |
| **QNAME**  | Query name |
| **FLAG** | Information describing the alignment (SAM/BAM) |
| **RNAME** | Reference sequence name, often contains the Chromosome name**. (=Chrom)** |
| **POS** | (Different information depending on the format used: where this alignment maps to the reference, POS. For SAM, the reference starts at 1, so this value is 1-based, while for BAM the reference starts at 0,so this value is 0-based.) |
| **MAPQ** | Maping Quality Map |
| **CIGAR** | String indicating alignment information that allows the storing of clipped |
| **MRNM/RNEXT** | the reference sequence name of the next alignment in this group, |
| **ISIZE/RNEXT** | length of this group from the leftmost position to the rightmost position, |
| **MPOS/RNEXT** | leftmost position of where the next alignment in this group maps to the reference, |
| **SEQ** | the query sequence for this alignment, |
| **QUAL** | leftmost position of where the next alignment in this group maps to the reference,  |
| **TAGs** | Additional optional information is also contained within the alignment |

 |

**Partial equivalence in Identification labels. Without equivalence in content**

**Full Equivalence:** **labels and content**

## Annex 2. Ontologies analyzed

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Title** | **Creator/Publisher** | **URL** |
| 1 | Gene Ontology | Genontology Consortium | <http://geneontology.org/> |
| 2 | Crop Ontology (CO): CGIAR Wheat Anatomy and Development Ontology | Generation Challenge Programme: cultivating plant diversity for the resource poor (CGIAR) | http://pantheon.generationcp.org/index.php?option=com\_content&task=view&id=472&Itemid=261 |
| 3 | Crop Ontology (CO): CGIAR Wheat Trait Ontology | Generation Challenge Programme: cultivating plant diversity for the resource poor (CGIAR) | <http://pantheon.generationcp.org/index.php?option=com_content&task=view&id=466&Itemid=261> |
| 4 | Plant Structure Development Stage Ontology. Plant Ontlogy (PO) | Plant Ontology Consortium | <http://www.plantontology.org/>  |
| 5 | Plant Anatomical Entity Ontology. Plant Ontlogy (PO) | Plant Ontology Consortium | <http://www.plantontology.org/>  |
| 6 | Trait Ontology (TO) | Plant Ontology Consortium | http://palea.cgrb.oregonstate.edu/trait\_ontology/cgi-bin/trait\_amigo/browse.cgi?session\_id=5611amigo1393472542; permalink: http://palea.cgrb.oregonstate.edu/trait\_ontology/cgi-bin/trait\_amigo/browse.cgi?action=permalink&open\_1=all |
| 7 | Sequence ontology (SO) |  Gene Ontology Consortium and GMOD (Generic Model Organism Database) | <http://www.sequenceontology.org/> |
| 8 | Project specific trait |  | There is not a specific source. It refers to multiple ontologies used for several trait projects.  |
| 9 | [ECPGR](http://www.ecpgr.cgiar.org/germplasm_databases/central_crop_databases/ECCDB_List.html?network=7) | Is refering to ECPGR database?Entire database or only European Wheat Database (EWDB): http://genbank.vurv.cz/ewdb/ | The source identified by scientists is a list of descriptors located in EWDB?It is a tab delimited format, data are downloded in ASCII |
| 10 | Ontologies to develop a conceptual ABM |  | There is not a specific ontology. It refers to a Agent-Based Model (ABM). |
| 11 | PATO (Phenotypic quality ontology), XEML | There is not information | http://obofoundry.org/wiki/index.php/PATO:Main\_Page#What\_is\_PATO.3F |
| 12 | Plant Environmental Condition Ontology (PECO) | Plant Ontology Consortium | <http://wiki.plantontology.org/index.php/Plant_Environment_Ontology_Wiki> |
| 13 | Plant pathogens | Lorenz Bülow /Institute of Genetics at the Technical University of Braunschweig, Germany | <http://www.pathoplant.de/> |
| 14 | PLEXdb (Plant Expression Database): WheatPLEX | PLEXdb Group | <http://www.plexdb.org/plex.php?database=Wheat> |
| 15 | QUDT Ontologies | Ralph Hodgson, TopQuadrant, Inc.; Paul J. Keller, NASA AMES Research Center; Jack Hodges and Jack Spivak | There is not a specific source. The QUDT Ontologies, and derived XML Vocabularies, are being developed by TopQuadrant and NASA. Originally, they were developed for the NASA Exploration Initiatives Ontology Models (NExIOM) project, a Constellation Program initiative at the AMES Research Center (ARC). See: http://qudt.org/ |

1. **Office** for **Partnerships**, Advocacy and Capacity Development at the Food and Agriculture Organization of the United Nations [↑](#footnote-ref-1)
2. <http://www.ciard.net/sites/default/files/RDA_Wheat_data_standards_survey_results.pdf> . Last accessed September 24th, 2014 [↑](#footnote-ref-2)
3. Doerr, M. 2003. The CINDOC Conceptual Reference Model: An Ontological Approach to Semantic Interoperability data. AI Magazine, V24(3), pp.75-92 [↑](#footnote-ref-3)
4. This work is in progress. So, information in this document could be change over time by Wheat Data Interoperability Working Group. [↑](#footnote-ref-4)
5. File Formats. Integrative Genomics Viewer. <http://www.broadinstitute.org/software/igv/home> . Last accessed September 24th, 2014 [↑](#footnote-ref-5)
6. GenBank format. Uril: . Url: <http://www.nmpdr.org/FIG/wiki/view.cgi/FIG/GBK> [↑](#footnote-ref-6)
7. Isa-tab. Url: <http://www.isa-tools.org/examples.html> [↑](#footnote-ref-7)
8. Germplasms identificators. Biodiversity Information Standards. Url: <http://wiki.tdwg.org/twiki/bin/view/GUID/GermplasmIdentificators>. [↑](#footnote-ref-8)
9. Alecrcia, A; Diulgheroff, S.; Mackay, M. FAO/Biodiversity multi-crop passport descriptors v2 (MCPD V2). FAO, June 2012, 11p. Url: <http://www.bioversityinternational.org/uploads/tx_news/FAO_Bioversity_Multi-Crop_Passport_Descriptors_V.2__MCPD_V.2_._1526_02.pdf> [↑](#footnote-ref-9)
10. Developing crop descriptors lists: guidelines for developers. Biodiversity International. Biodiversity Technical Bulletin, N.13. [↑](#footnote-ref-10)
11. Analysis of Survey: data standards in the wheat research community. Wheat Data Interoperability WG, 07/04/2014-03/06/2014 [↑](#footnote-ref-11)
12. Wang, R. and Strong, D.M. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems, 1996, Vol.12 (4), pp.5-34 [↑](#footnote-ref-12)
13. Notes: A linked vocabulary suppose more richness for data content [↑](#footnote-ref-13)
14. Reverté, C. 2014. Datasets interoperability assessment: CIARD RING case. CIARD-RING, 20/09/2014. Url: <http://ring.ciard.net/sites/default/files/Datasets-Interoperability-Assessment-CIARD-RING-ver1.pdf> [↑](#footnote-ref-14)
15. If the attribute is empty the value is always 0. More attributes an indicator has, more quality it provides. [↑](#footnote-ref-15)
16. Scores are only examples of the methodology applied. The values can be changed following the same method. The total value represents the result of the sum of the values. [↑](#footnote-ref-16)
17. It is given a highest value because they are interoperable by definition. [↑](#footnote-ref-17)
18. Accessibility: in some cases is possible to browse in the ontology using information systems but in other cases they use to be integrated in repositories of ontologies or in databases such as GO Database trought the use of AMIGO browse and search engine system: <http://amigo2.berkeleybop.org/amigo> [↑](#footnote-ref-18)
19. Mappings made in OBO. Url: <http://www.obofoundry.org/index.cgi?show=mappings> [↑](#footnote-ref-19)
20. Gene Ontology Annotations and Resources. The Gene Ontology Consortium, 2012. *Nucleic Acids Research*, 2012, 1-6. Url. <http://nar.oxfordjournals.org/content/early/2012/11/16/nar.gks1050.full.pdf>

 [↑](#footnote-ref-20)