

**RESEARCH ARTICLE****COMMUNICATION AMONG MEDICAL INFORMATION SYSTEMS**  
principles for elaborating definitions

Amanda Damasceno de Souza<sup>1</sup>  <https://orcid.org/0000-0001-6859-4333>

Maurício Barcellos Almeida<sup>2</sup>  <https://orcid.org/0000-0002-6153-1011>

<sup>1</sup> Federal University of Minas Gerais, Belo Horizonte, MG, Brazil / e-mail: [amanda@ufmg.br](mailto:amanda@ufmg.br)

<sup>2</sup> Federal University of Minas Gerais, Belo Horizonte, MG, Brazil / e-mail: [mba@eci.ufmg.br](mailto:mba@eci.ufmg.br)

**ABSTRACT**

The formulation of definitions for medical terms to be applied in ontology-based information systems is a fundamental activity for supporting continuing care, in cases in which a patient moves among different medical units. However, the practice of creating good definitions is not a trivial task, and the literature does not provide clear methodologies. This article aims to present a method, we called *OntoDef*, to establish principles for systematizing the process of creating standardized definitions, which can be capable to provide the expected communication between the medical information systems and, therefore, to foster better services for citizens. The creation and testing of such method was conducted in the field of leukemias in the scope of the Blood Ontology project. The results show that definitions of biomedical entities are defined in a variety of criteria, which do not always meet the ontological representation requirements for adequate communication between medical information systems. Some well-known issues already identified are the identification of essential features of an entity, in addition to involve circularity, intangibility as well as the complexity in diagnosing manifold diseases such leukemias. We hope this experience can contribute to improvements in research projects in which Information Science supports Medicine and Healthcare with the aim of providing a better service to the citizen.

**KEYWORDS**

Knowledge organization systems. Ontologies. Concepts. Medical sciences.

**COMUNICAÇÃO ENTRE SISTEMAS DE INFORMAÇÃO MÉDICOS**  
princípios para elaborar definições**RESUMO**

A formulação de definições de termos médicos para uso em sistemas de informação baseados em ontologias é fundamental para a continuidade do cuidado quando um paciente se move entre diferentes unidades médicas. Entretanto, a prática de criar boas definições não é uma tarefa trivial e a literatura não contempla metodologias para tal. Este artigo tem como objetivo apresentar um método, chamado *OntoDef*, que estabelece princípios para sistematização do processo de criação de definições padronizadas, as quais possam proporcionar uma melhor comunicação entre sistemas de informação médicos, e, portanto, melhor atendimento ao cidadão. A criação e teste do método foi conduzido no domínio das leucemias, no âmbito do projeto *Blood Ontology*. Os resultados mostram

que definições de entidades biomédicas são definidas por uma variedade de critérios, que nem sempre atendem aos requisitos de representação ontológica para a adequada comunicação entre sistemas médicos. Problemas verificados dizem respeito a identificação de características essenciais de cada entidade e envolvem a circularidade, a intangibilidade e a complexidade em diagnosticar doenças complexas como leucemias. Espera-se que essa experiência possa contribuir para melhorias nos projetos em que a Ciência da Informação apoia a medicina e os cuidados à saúde na busca por melhor atendimento ao cidadão.

**PALAVRAS-CHAVE:**

Sistemas de organização do conhecimento. Ontologias. Conceitos. Medicina.



**JITA** ID. Knowledge representation.

## 1 INTRODUCTION

The communication between health facilities is essential for the enduring healthcare, for example, in long-term cancer treatments and prenatal care. Such communication is currently performed mainly via information systems, which usually are not able to interact properly because of several factors. One of these factors involves conceptual differences between terms and their definitions. In the context of ontological studies, well-founded representations of the medical field are created exactly to improve communication across different geographic and temporal scopes.

Ontologies in the biomedical domain are relevant artifacts that allow to connect common elements between data from biology and related areas. Biomedical ontologies are logically structured in a hierarchical way, connecting classes to represent entities and their relationships. In this context, the Open Biological and Biomedical Ontologies (OBO) Foundry (2018) was created to provide principles that help in the standardization of ontologies for the biomedical domain (MEEHAN et al., 2011). According to Seppälä and Ruttenberg (2013), ontologies based on OBO Foundry principles include both textual (natural language) and formal (logical) definitions. Definitions are very important to standardize and represent data, vocabularies and terminologies, as well as to create taxonomies, that ultimately seek to promote communication between systems (KEIZER; ABU-HANNA; ZWETSLOOT-SCHONK, 2000).

Specifically, within ontologies underlying modern medical information systems, definitions of terms play an important role, serving a number of purposes: they enable hierarchies to be created, they validate class and category arrangements, they clarify specialized terms allowing to be understood and data to be integrated (IWOOD, 2014). In Information Science (IS), the classical theories on definitions were advocated by the German Philosopher Ingetraut Dahlberg through her **Theory of the Concept**, as an important phase in the construction of Thesaurus (DAHLBERG, 1978a, b). For Dahlberg (1978a, p.106), definitions play an important role in the formulation of concepts, since "they are indispensable presuppositions in argumentation and verbal communications and are necessary elements in the construction of scientific systems".

However, the question is how to formulate these definitions in line with information systems aligned with modern digital environments (KÖHLER et al., 2006). The difficulty in creating definitions in ontologies is related to the fact that, besides the complexity and cost of the task, it is susceptible to errors and needs to be performed by trained people (TSATSARONIS et al., 2013). The ontology includes necessary features to solve problems concerning the definition of terms (SPEAR, 2006). However, formulating definitions requires specific knowledge of ontological theories and requirements, including the point of view of Medicine and Information Science.

In order to write definitions with well-formed criteria, one should start with the definition in natural language, as suggested by Gruber (1993). In the sequel, she needs to formulate definitions in a logical representation language, reaching then the definitions used by intelligent information systems of the Semantic Web age. The methodologies for building ontologies as artifacts suggest the presence of definitions, without explaining, however, how they should be formulated. This article is an initiative in this sense in the field of Information Science. Ways to standardize definitions is essential for the build them rigorous and concise, which allows them to be used to locate the term in the taxonomic structure of an ontology.

The remainder of this article has been organized as follows: Section 2 provides a brief introduction on definitions and their characteristics; Section 3 introduces the leukemia domain and presents the systematized method applied to terms of the leukemia domain; section 4

discusses the results and section 5 makes brief concluding remarks. It is worthwhile to note that since the present research is part of an international project, some parts of the article are left in their original foreign language, without prejudice to medical understanding.

## 2 DEFINITION FEATURES

Relevant criteria for the ontology project are clarity and coherence: "all definitions must be documented in natural language (GRUBER, 1993, p.2 and 3). Also, Uschold (1996, p.12 and 13) presents several basic criteria for the creation of definitions, among them clarity and coherence, as cited below:

a) Clarity: definitions must be clear and unambiguous when expressed in natural language or formally codified. It is recommended to use examples to illustrate what is intended as well as negative examples to make clear what is not intended;

b) Consistency and coherence: an ontology must be internally and externally consistent; circularity must be avoided; new terms must be adopted only after being clarified;

c) Extensibility and re-usability: ontology should be designed in a way that allows its re-use and extensibility, and for this, definitions should be well formed.

Even though such recommendations are present in the literature, several problems persist when defining terms of complex fields like medicine. Seppälä and Ruttenberg (2013) conducted a study on practices for creating definitions in ontologies in which it is reported main dimensions of definitions issues (Table 1).

| 4

**Table 1.** Problems with settings

Content	Logic	Writing and style	Coverage
Insufficient information, very informative or very complex, outdated or obsolete, lack of standard	Imprecise or vague, circular, contradictory.	Poorly worded, inconsistent style.	Several similar definitions, absence of definition.

Source: adapted from Seppala and Rutemberg (2013).

Table 2 illustrates the definition problems of Gene Ontology (GO)<sup>1</sup> and NCI-Thesaurus<sup>2</sup>:

<sup>1</sup> <http://geneontology.org/>

<sup>2</sup> <https://ncithesaurus.nci.nih.gov/ncitbrowser/>

**Table 2.** Examples of Definitions Problems

Content	Logic	Coverage
<p>GO:0050566 <b>term:</b> asparagine-tRNA synthase (glutamine-hydrolyzing) activity. <b>definition:</b> Catalysis Cyc:6.3.5.6-RXN</p>	<p>GO:0042270 <b>term:</b> Protection from natural killer cell mediated cytotoxicity <b>definition:</b> The process of protecting a cell from cytotoxicity by natural killer cells</p>	<p><b>NCI Thesaurus</b> <b>term:</b> Acute Myeloid Leukemia. <b>definition:</b> A clonal expansion of myeloid blasts in the bone marrow, blood or other tissues. The classification of acute myeloid leukemias (AMLs) encompasses four major categories: 1) AML with recurrent genetic abnormalities; 2) AML with multilineage dysplasia; 3) Therapy-related AML; 4) AML not otherwise specified. The required bone marrow or peripheral blood blast percentage for the diagnosis of AML is 20% (WHO classification). <b>Other definitions:</b> An aggressive (fast-growing) disease in which too many myeloblasts (immature white blood cells that are not lymphoblasts) are found in the bone marrow and blood.  A rapidly progressive cancer of the blood and bone marrow consisting of the proliferation of abnormal myeloblasts, which are immature, dysfunctional white blood cells.</p>
<p>GO:0001655 <b>term:</b> urogenital system development. <b>definition:</b> the development of the urogenital system</p>	<p>GO:0006190 <b>term:</b> inosine salvage. <b>definition:</b> Any process that generates inosine, hypoxanthine riboside, from derivatives of it without de novo synthesis.</p>	<p><b>NCI Thesaurus</b> <b>term:</b> Chronic Myelogenous Leukemia. <b>definition:</b> A chronic myeloproliferative neoplasm characterized by the expression of the BCR-ABL1 fusion gene. It presents with neutrophilic leukocytosis. It can appear at any age, but it mostly affects middle aged and older individuals. Patients usually present with fatigue, weight loss, anemia, night sweats, and splenomegaly. If untreated, it follows a biphasic or triphasic natural course; an initial indolent chronic phase which is followed by an accelerated phase, a blast phase, or both. Allogeneic stem cell</p>

		transplantation and tyrosine kinase inhibitors delay disease progression and prolong overall survival. <b>Another definition:</b> A slowly progressing disease in which too many white blood cells are made in the bone marrow.
<p><b>NCI Thesaurus term: Tuberculosis</b></p> <p><b>definition:</b> A chronic, recurrent infection caused by the bacterium <i>Mycobacterium tuberculosis</i>. Tuberculosis (TB) may affect almost any tissue or organ of the body with the lungs being the most common site of infection. The clinical stages of TB are primary or initial infection, latent or dormant infection, and recrudescence or adult-type TB. Ninety to 95% of primary TB infections may go unrecognized. Histopathologically, tissue lesions consist of granulomas which usually undergo central caseation necrosis. Local symptoms of TB vary according to the part affected; acute symptoms include hectic fever, sweats, and emaciation; serious complications include granulomatous erosion of pulmonary bronchi associated with hemoptysis. If untreated, progressive TB may be associated with a high degree of mortality. This infection is frequently observed in immunocompromised individuals with AIDS or a history of illicit IV drug use.</p>	<p><b>Term: Tuberculosis</b></p> <p><b>definition:</b> A chronic, recurrent infection caused by the bacterium <i>Mycobacterium tuberculosis</i>.</p>	<p><b>NCI Thesaurus term: Acute Leukemia.</b></p> <p><b>definition:</b> A clonal (malignant) hematopoietic disorder with an acute onset, affecting the bone marrow and the peripheral blood. The malignant cells show minimal differentiation and are called blasts, either myeloid blasts (myeloblasts) or lymphoid blasts (lymphoblasts).</p> <p><b>Another definition:</b> A rapidly progressing cancer that starts in blood-forming tissue such as the bone marrow and causes large numbers of white blood cells to be produced and enter the blood stream.</p>

Source: Köhler *et al.* (2006, p.9), National Cancer Institute Thesaurus (2020) e Smith (2013).

Spear (2006) explains that if one wants that separate information systems storing biomedical information to be compatible or interoperable, consensual terminology and consistent definitions must be applied. Thus, the terms need to be internally coherent, supported and aligned with facts of the reality of biomedical sciences. Thus, basic principles need to be adopted for the formulation of appropriate definitions, such as those cited in the remainder of this section.

*2.1 Principle of necessary and sufficient conditions*

Definitions created according to Aristotelian principles emphasize necessary and sufficient conditions. Indeed, a definition is a declaration of necessary and sufficient conditions (SMITH, 2013). For example: "to be an entity A is a necessary condition to be an entity, that is, all B is an A; to be an entity A is a sufficient condition to be an entity B, that is, all A is a B; [...]"

if these two statements remain, to define an A is to define something that satisfies a B [...]. This is because A is a more difficult term to understand than B" (SMITH, 2013, p.1).

Swartz (2010) explains the relationship that holds between "is a sufficient condition" and "is a necessary condition" is the inverse, that is, if x is a sufficient condition for y, then y is a necessary condition for x. In this statement, the implication of the inverse of a statement is the result of the inverse of its two parts.

## 2.2 Principle of single inheritance

The role of definitions in an ontology is to specify how to exhibit the characteristics of terms representative of entities in a consistent way, in such a way that transitive inheritance through the hierarchy is ensured. Consistency in definitions requires a unifying point of view – the context – in the representation, which must be maintained throughout the ontology. In meeting these requirements, the position of a term in the hierarchy will be determined by its own definition and by the definition of the term above.

The inheritance principle contains two relevant aspects of Aristotelian definitions: *Genus* and *Differentia*. The *Genus* is the essential characteristic that all of us underground the hierarchy will assume, that is, an existing definition contributes to the part of a new definition; the *Differentia*, on the other hand, seeks to define the essential characteristics that distinguish a type from other types, that is, it assigns a unique characteristic to the term (MICHAEL; MEJINO JUNIOR; ROSSE, 2001, p.463).

According to Seppälä, Schreiber and Ruttenberg (2014), the first part of the definition is the *Genus* part and the second part is the *Differentia* part. Definitions should be formulated following the structure of *genus plus differentia* (SEPPÄLÄ, 2016). Multiple inheritance occurs when an entity has two parents in the hierarchy. By avoiding multiple inheritance, ambiguities and issues of understanding are also avoided.

| 7

## 2.3 Principle of non-circularity

Köhler et al. (2006, p.2) recommend rules based also on Aristotelian principles to formulate definitions in ontologies, among them avoiding circularity. Circularity occurs when a term is used to define itself, for example, wireless telephone is a telephone that has no wire.

Attention is therefore due to the circularity rule to avoid that a definition uses its own terms to define itself. In ontology, it is important to follow principles of non-circularity, to enable definitions written clearly according to the real meaning of the term. Non-circularity provides concise definitions, in short, complete sentences (KÖHLER et al., 2006).

## 2.4 Principle of intangibility

When formulating definitions in ontologies, the so-called "figurative or obscure language" should be avoided, that is, a definition should clarify the meaning of an unknown word and not hinder its understanding. This rule is called the principle of intangibility, which suggests that the definition should use only intelligible terms, easily understood, since a definition should be understood by persons who are not field specialists (KÖHLER et al., 2006).

The principle of intangibility refers to the quality of terminology, so that each defined term must meet basic standards of understanding. Thus, the definition is intelligible and fully comprehensible without the need for prior reading or additional consultation to information sources. This also includes avoiding the use of technical jargon (KÖHLER et al. 2006).

### 3 DEFINITIONS IN THE FIELD OF LEUKEMIAS

To propose a method to formulate definitions in natural language appropriate to biomedical information systems ontologies, we studied the case of Acute Myeloid Leukemia (AML) within the Blood Ontology (BLO) project (ALMEIDA et al., 2013). BLO is a formal vocabulary covering knowledge in hematology and hemotherapy, which aims to gather together, organize and facilitate the manipulation of data on human blood. The ontology describes hematological neoplasms including leukemias and lymphomas. In the present paper, our goal is to define the terms of the leukemias present in the BLO.

When defining a domain, an important step is the knowledge acquisition phase, which makes it possible to know about the domain. This phase allowed us to understand that leukemias are complex diseases and, therefore, in the last 30 or 40 years, cytogenetic analysis of hematological neoplasms has been one of the hematology fields that has most developed (NAJFELD, 2009). Leukemias are classified according to the speed of cell differentiations into: i) acute, in which abnormal cells are characterized by rapid development; ii) chronic, in which abnormal cells develop more slowly. Leukemias are further differentiated by the type of white blood cell affected, according two types: iii) myeloid, which starts in the myeloid cells, also called myeloblastic; iv) lymphoid, which starts in the lymphoid cells and is known as lymphoid, lymphoblastic or lymphocytic leukemia (NATIONAL CANCER INSTITUTE, NCI, 2013b).

AML thus refers to a group of heterogeneous diseases with respect to clonality, chromosomal changes and response to treatment. It is a disease that exhibits a i) phenotypic heterogeneity, that is, it refers to the real properties observed in an organism, such as morphology, development or behavior; and ii) genotypic heterogeneity, that is, the one relative to the hereditary genetics of the individual. It is characterized by proliferation and accumulation of immature hematopoietic cells from bone marrow and blood. These malignant cells gradually replace normal erythroid precursors, myeloid and megakaryocytes and inhibit growth and maturation. Its definition of AMLs is based on the diagnosis established by morphology, immunophenotypic, cytogenetic and molecular criteria (ARBER; COUSAR, 2013). Blastic cells usually make up 1% to 5% of marrow cells. When the number of blasts is above 20% – a criteria defined by the World Health Organization (WHO) – the diagnosis is AML. This medical knowledge is necessary for the Information Science professional in order that she can draw up definitions in the field of leukemias.

### 4 SYSTEMATIZATION OF DEFINITIONS IN BIOMEDICAL ONTOLOGIES

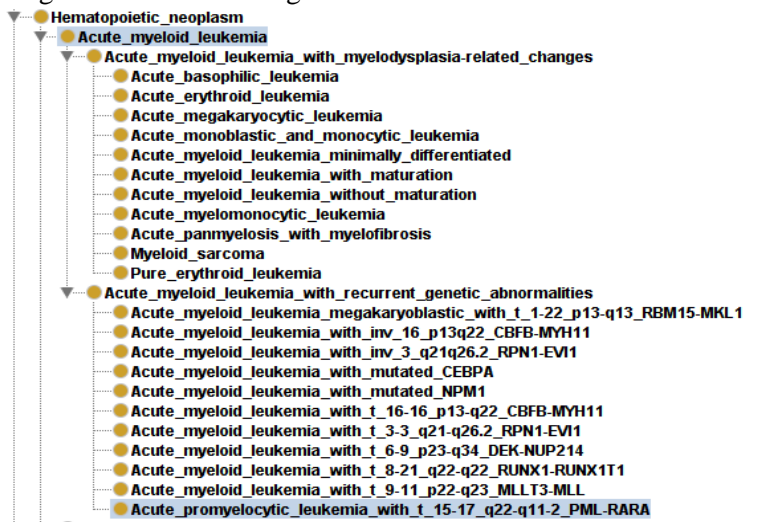
The steps in the method of formulating definitions in the field of AMLs are presented in the remainder of this section.



4.1 Sample for Survey

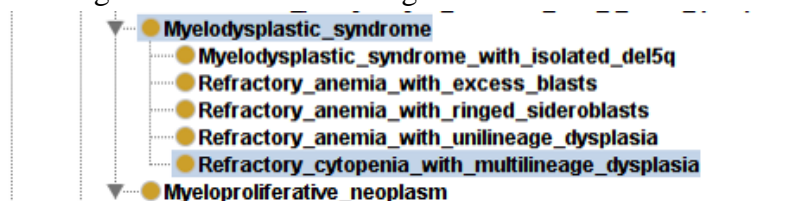
The cutout of the AML sample was defined by the set of terms that included the hematological neoplasms of the BLO, restricted to the analysis of leukemias that had the myeloid lineage as their progenitor and divided into three classes: i) for the AML class, there are a total of 25 terms; ii) for myelodysplastic syndrome, there are 6 terms, and iii) for myeloproliferative neoplasms, there are 12 terms. The total to be defined adds up to 43 terms, as shown in figures 1, 2 and 3, respectively. This is a convenience sample and the criterion adopted for extracting data favored terms from the part of the BLO about cancer, so that a total of 43 terms cover blood neoplasms. The main focus of the research was the first class, the AMLs, but in the class of myeloproliferative neoplasms, chronic leukemias were also defined.

**Figure 1.** Fragment of the first range to be defined with 25 classes



Source: Fragment of BLO in Protégé - Almeida et al., (2013).

**Figure 2.** Fragment of the second range to be defined with 6 classes



Source: Fragment of BLO in Protégé - Almeida et al., (2013).

**Figure 3.** Fragment of the third range to be defined with 12 classes



Source: Fragment of BLO in Protégé - Almeida et al., (2013).

A practical example of how to formulate a definition for a sample term is given in section 4.4: *Acute Myeloid Leukemia*.

#### 4.2 Acquisition of knowledge for definitions

In order to acquire knowledge, it is important to retrieve literature, texts and definitions of the domain in question from specialized sources. For the proposed method, definitions of the leukemia cited in the research sample were searched in the following online sources: NCI Thesaurus (NCIt); NCI Dictionary of Cancer Terms; Medical Subject Headings (MeSH); Medscape, Ontobee; Disease Ontology (DO); and in the text source Pathology of Foucar et al., (2012).

In this study, we sought health information sources that had cancer as a theme, some works indicated by experts in the oncology field, besides material from BLO. Another important step in knowledge acquisition is to reuse knowledge from other ontologies, such as Disease Ontology and Gene Ontology (XIANG et al., 2011; SCHRIML et al., 2012).

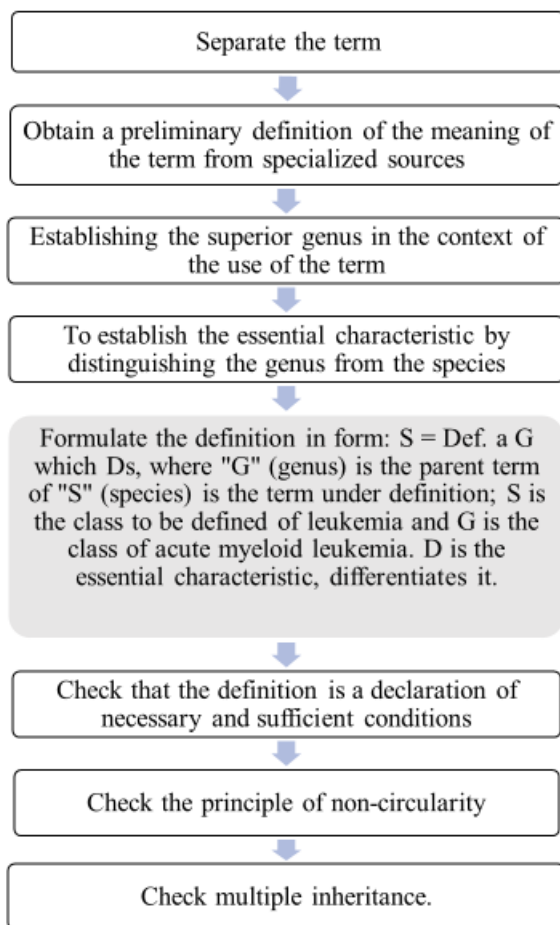
In the end, one needs to evaluate the existing definitions in the literature and thus verifies whether they present some of the problems already mentioned as described by Seppälä and Ruttenberg (2013).

#### 4.3 List of basic principles for creating definitions in natural language

When formulating definitions in natural language, the principles of non-circularity and tangibility we have already mentioned, as suggested by Köhler et al. (2006), must be followed. In order to provide definitions that are useful for understanding, a clear and precise way of use language is adopted, avoiding using terms inherent to the very term to be defined, and avoiding figurative or obscure language. When using technical terms from the biomedical field, one should describe it as accessible as possible.

In order to systematize the research method, the steps were based on good practices identified in the literature by various researchers such as Michael, Mejino Junior and Rosse (2001); Köhler et al. (2006); Smith et al. (2005); Soergel (2014); Seppälä and Ruttenberg (2013); Petrova et al. (2015); Tsatsaronis et al. (2013); Seppälä, Schreiber and Ruttenberg (2014); among others. The steps of the method are presented in Figure 4:

**Figure 4.** Method of drawing up definitions



Source: Research data.

After systematizing the basic principles to create definitions in natural language, we need to apply such principles to each term in the research sample. The following presents the result of the knowledge acquired in carrying out this stage of the research.

4.4 How to apply the definitions method

The following describes in more detail how to apply each step of the list of basic principles to create textual definitions (Table 3):

**Table 3.** List of basic principles for creating textual definitions

1) To separate the term	List the term to be defined. In the medical field, there may be different terms or synonyms for an illness.
2) To obtain a preliminary definition	Use domain-specific sources of information: books, articles, dictionaries, encyclopedias, thesaurus, ontologies.
3) To establish the superior <i>genus</i> in the context of the use of the term	The <i>genus</i> , the common characteristic to the terms of the hierarchy, defined by the ratio <is-a>, e.g. acute myeloid leukemia <b>is-a</b> hematopoietic neoplasm.

4) To establish the essential characteristic by distinguishing the <i>genus</i> from the species	Use differentiate to define the unique and essential characteristic (or group) that distinguishes one term from another. This essential characteristic usually requires an expert.
5) To formulate the first version of the definition in the form	<b>S = a G which Ds, where "G" (for: <i>genus</i>) is the parent term for "S"; "S" (for: species) where S is the leukemia class to be defined, G is the most general class and D is the differentiator that distinguishes the instance of S from the instance of D.</b>
6) To check necessary and sufficient conditions	Use the expression: to be an A is a necessary condition to be a B, so each B is an A. Being an A is a sufficient condition to be a B, so each A is a B; this means that "A" represents the essential characteristic and "B" represents the defined term.
7) To check the principle of non-circularity	Check whether the definition has used the term(s) itself to define itself.
8) To check presence of multiple inheritance	Check for the presence of hierarchy nodes that have two parent nodes.

Source: Research Data.

The following example is a real practical example of the method application for drawing up definitions in the field of leukemia (Table 4).

**Table 4.** Elaboration of the definitions of Leukemias

Stage	Example	Comment
1) To separate the term	<b>Term: Acute Myeloid Leukemia</b>	Leukemia term to be defined according to the BLO
2) To obtain a preliminary definition of the meaning of the term from expert sources	<b>NCI Thesaurus (NCIt): a clonal expansion of myeloid blasts in the bone marrow, blood or other tissues. The classification of acute myeloid leukemias (AMLs) encompasses four major categories: 1) AML with recurrent genetic abnormalities; 2) AML with multilineage dysplasia; 3) Therapy-related AML; 4) AML not otherwise categorized. The required bone marrow or peripheral blood blast percentage for the diagnosis of AML is 20% (WHO classification).</b>	The selected sources of information were: Disease Ontology, Ontobee, MeSH, Medscape, Book Foucar et al. (2012), NCI-Dictionary of Cancer Terms, NCI-Thesaurus.
3) To establish the superior <i>genus</i> in the context of the use of the term	<b>Acute myeloid leukemia is a hematopoietic neoplasm.</b>	In the case of leukemias, the genus common to the terms is acute myeloid leukemia which has a characteristic abnormal shunt in the myeloid lineage. The <is-a> ratio was used to establish the <i>genus</i> .

<p>4) To establish the essential characteristic by distinguishing the <i>genus</i> from the species</p>	<p><b>Derives from an uncontrolled and rapidly proliferation of the myeloid lineage and their precursors.</b></p>	<p>The fact of simply being of the myeloid lineage does not characterize the LMA, but the change in the myeloid lineage, because there is also the differentiation to the lymphoid lineage. In the case of LMA, the differentiation between leukemia was defined by the type of cell involved (differentiation from the myeloid lineage) and the percentage of immature cells (blasts).</p>
<p>5) To formulate the first version of the definition in form: S = Def. a G which Ds, where "G" (for: <i>genus</i>) is the parent term of "S"; "S" (for: species) where S is the class to be defined, G is the most general class and D is the <i>differentia</i> that characterizes the instance of S from the instance of D</p>	<p><b>An Acute myeloid leukemia is-a hematopoietic neoplasm that derives from an uncontrolled and rapidly proliferation of the myeloid lineage and their precursors.</b></p>	<p>Avoiding the use of unusual words, for example, in the definition of Myeloid neoplasms with abnormalities of PDGFRA PDGFRB or FGFR1, has been changed aberrantly by abnormal.</p>
<p>6) To check necessary and sufficient conditions</p>	<p>Being an LMA is sufficient condition to "derive from an uncontrolled and rapidly proliferation of a myeloid lineage and its precursors", that is, each LMA derives from an uncontrolled and rapidly proliferation of a myeloid lineage and its precursors.</p>	<p>Perform the single inheritance check to confirm that the characteristic defined in item 4 is necessary and sufficient for each type of leukemia and that there are other leukemias or other diseases with these same characteristics.</p>
<p>7) To check the principle of non-circularity</p>	<p>There seems to be no circularity, we keep the definition as in (5).</p>	<p>To correct the circularity of more complex terms, it is sometimes necessary to describe the meanings of the acronyms in full and to clarify the form of mutation: whether it was translocation, deletion or inversion. The most difficult term to avoid circularity was Acute myeloid leukemia with mutated CEBPA.</p>

8) To check multiple inheritance	Unique inheritance: myeloid lineage.	Because it's a clonal disease, it could have myeloid or lymphoid lineage. Thus, myeloid inheritance was defined according to the classification criteria of FAB/WHO (2008). However, Silva et al. (2006) reports the existence of types of acute leukemia with both lineages (myeloid and lymphoid), being considered mixed, hybrid or biphenotypic leukemias.
----------------------------------	--------------------------------------	--

Source: Research data.

The next section, Section 4.5, presents the validation that has been performed by a cancer domain expert in formulating the sample definitions.

4.5 Validation of settings with expert

The validation of knowledge inserted in the definitions was performed on the basis of the necessary and sufficient conditions, with the help of a medical expert, namely, an oncologist. This process has been previously tested in the BLO project and consists of showing to the expert possibilities of definition, then requesting validation and approval (COELHO, 2012).

For the validation of the definition, the physician (domain specialist) was asked:

- i) Is would be correct to state that any entity deriving from an uncontrolled proliferation of a Myeloid strain and its precursors is the essential characteristic of an Acute Myeloid Leukemia?
- ii) Are there other entities deriving from an uncontrolled proliferation that are not Acute Myeloid Leukemia?

In this context the oncologist stressed:

“To define an AML, it is necessary to observe three characteristics: morphological, cytogenetic and immunophenotype. The AML is of unique origin, the myeloid lineage, as it is a clonal disease and thus descends from only one lineage (the myeloid). Leukemia, when it is biphenotypic or bi-line, has two origins because it presents several populations of clonal cells... however, this type presents only the myeloid lineage. Even if the offspring of the myeloid lineage is minimal, its presence suggests differentiation. The number of leukemic cells grows rapidly and disease worsens in a short period of time.”

The conclusion is that the essence of entity was correctly captured by the definition and it was decided to maintain the definition as formulated by the method: “An Acute myeloid leukemia is a hematopoietic neoplasm that derives from an uncontrolled and rapidly proliferation of the myeloid lineage and their precursors”. A similar process was repeated for the other entities that separated for definition purposes.

## 4 DISCUSSION

In order to create good definitions, one firstly needs to seek sources of medical information. However, even if she is an expert in medicine, she is not specialist in classification as a Librarian. When searching for sources of knowledge, it is necessary to analyze the quality criteria as suggested by Uschold (1996, p.12 and 13), Gruber (1993, p.2 and 3), Seppälä and Ruttenberg (2013), Seppälä, Schreiber and Ruttenberg (2014), Köhler et al. (2006) and Spear (2006).

Several specialized dictionaries were used to obtain definitions of leukemia due to the fact that some definitions are not very informative or do not have essential characteristics. In the case of leukemias, it was necessary to use more sources of definitions, already cited, such as the pathology books of Foucar et al. (2012), the hematology books of Hoffman et al. (2008), the classifications of leukemias from FAB and WHO, scientific articles, among other sources. Several health information sources were consulted for the research in order to analyze the existing definitions.

The textual definitions of the sources investigated reveal several issues described by Seppälä and Ruttenberg (2013), such as: circular, intangible, use of technical terms, as well as multiple definitions for the same term. In this research, for example, the first term of the sample – acute myeloid leukemia with myelodysplasia-related changes – revealed circularity; the second term of the sample – acute myeloid leukemia with recurrent genetic abnormalities – showed intangibility, due to the presence of technical terms. In order to soften intangibility, it was necessary to explain the meaning of genetic mutations to better understand the definition of leukemia.

Validation with the cancer expert was essential to define the essential characteristic of each AML class, as the oncologist has knowledge to define each type of leukemia. The field expert, in analyzing what was the necessary and sufficient condition of each leukemia class used the medical literature to confirm the essential characteristic, especially when there were more characteristics. Validation was precisely the interference of the expert that determined the essence of an entity for definition, which was confirmed based on the FAB and WHO diagnostic criteria. The last review of the diagnosis of leukemia was performed by the WHO, in 2008, when it brought together specialists to review the definitions of the AMLs. However, some types were not reviewed due to rarity and lack of consensus on the diagnosis.

The essence was defined mainly by morphological characteristics, except for the term acute myeloid leukemia with recurrent genetic abnormalities, in which the essence was based on the characteristics of cytogenetic abnormalities. This was a term whose definition proved to be still more complex, since it has characteristics that may be present in other types of leukemia, in addition to present circular and intangible definitions.

## 5 FINAL CONSIDERATIONS

Defining terms in the leukemia field has proved to be an arduous, complex and time-consuming task. The terminological complexity of leukemia has made the work even more laborious, considering the application of the method in a context without consensus in the definition of leukemia, besides requiring several scientific perspectives for diagnosis. The main difficulty in defining terms was to find the essential characteristic of leukemia and other terms, since the field itself is one of the fields studying cancer that presents the greatest diversity of

phenotypic and genetic alterations at the time of diagnosis.

Indeed, cancer is a complex disease, especially in the case of leukemias. Nowadays, only one morphological characteristic – the presence of 20% blasts – is not sufficient for the diagnosis and treatment of the disease. There are several characteristics analyzed to define leukemia. In fact, AMLs are considered one of the most complex cancers in relation to diagnosis and treatment, presenting high mortality rates. The FAB and WHO classifications still presented difficulties when categorizing, defining and diagnosing the subtypes of AML. Efforts to better categorize myeloid neoplasms are many, as Vardinan, Harry and Brunning (2002) point out, who conducted a review of previous classifications of myeloid neoplasms and acute leukemia. Vardiman et al. (2009) published a study on updating the WHO classification, which in addition to using morphology to define leukemias, added information on immunophenotyping, cytochemistry, genetics and clinical features. All this added complexity to the search for the essence of each class of leukemia, considering also that science is currently studying ways to better categorize this field.

Having an overview of the domain to be defined is fundamental to apply the method of definitions. In the case of leukemia, it was necessary to study the pathology, diagnosis, terminology, and etiology, in addition to discuss all this information with the oncologist, reaffirming the importance of knowing the domain.

The method is applicable to other fields of medicine and one of the main contributions of this research, besides the practical example of how to elaborate well-formed definitions, is the systematization effort to define the essential characteristic of each type of leukemia. As future work, we hope to automate the creation of formal definitions according to the BFO guidelines.

It is the responsibility of this information professional, who works in the health area, to develop research on themes that involve ontologies, such as representation tools for information systems, which can contribute to standardization, interoperability among medical data, and communication among health institutions. It is important to emphasize that the work of the researcher as a Librarian, with the medical team in the cancer area, was essential for the development of the method we called *OntoDef*. The experience contributed to the acquisition of knowledge about leukemia, as well as to the validation of textual definitions.

In this context, the Librarian and the information professional have the challenge of contributing to the development of biomedical ontologies, based on their expertise in the construction of thesaurus, classifications, terminologies, and their ability to analyze subjects and evaluate sources of information. Well-built definitions, based on universal principles, are the first step toward achieving communication and continuity of proper care for citizens, given the current prevalence of information systems.

## REFERENCES

ALMEIDA, Maurício Barcellos *et al.* A ontologia do sangue: uma iniciativa para representação e organização do conhecimento sobre Hematologia e Hemoterapia. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro, v.7, n.1, mar., 2013.

ALMEIDA, Maurício Barcellos *et al.* The Blood Ontology: an ontology in the domain of hematology. Status. In: INTERNATIONAL CONFERENCE OF BIOMEDICAL ONTOLOGIES, 2011, US. **Proceedings of the...** Available on: <http://ceur-ws.org/Vol-833/>. Access at: abr. 2014.



ARBER, Daniel A.; COUSAR, John B. Hematopoietic Tumors: Principles of Pathologic Diagnosis. Chapter 71. *In: GREER, John P. et al. Wintrobe's clinical hematology*. 13th. City: publication, 2013.

COELHO, Katia Cardoso. **Aquisição de conhecimento especializado para construção de ontologias**: um estudo no domínio das ciências da vida. 2012.174f. Dissertação (Mestrado em Ciência da Informação). Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2012. Availale on:

[http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/ECID-8XLQS4/disserta\\_o\\_katia\\_cardoso\\_coelho.pdf?sequence=2](http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/ECID-8XLQS4/disserta_o_katia_cardoso_coelho.pdf?sequence=2). Access at: ago. 2014.

DAHLBERG, Ingtraut. Teoria do Conceito. **Ci. Inf.**, Rio de Janeiro, v.7, n.2, p. 101-107, 1978a. (Trad. Astério Tavares Campos).

DAHLBERG, Ingtraut. A Referent-oriented analytical concept theory of interconcept. **International Classification**, v.5, n.3, p.142-150, 1978b.

DISEASE ONTOLOGY (DO). **Leukemia**. 2014. Availale on: <http://disease-ontology.org/>. Access at: 29 set 2014.

FOUCAR, Kathryn *et al.* Myeloid neoplasm. *In: FOUCAR, Kathryn.; et al. Diagnostic pathology blood and bone marrow*. Monitoba:Amirsys, 2012. cap.9. 2-208.

GRUBER, Thomas. R. **Toward principles for the design of ontologies used for knowledge sharing**. Revision: August 23, 1993. Stanford Knowledge Systems Laboratory. Availale on: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.3207&rep=rep1&type=pdf>. Access at: 30 out 2013.

HOFFMAN, Ronald; et al. **Hematology**: basic principles and practice, 5th Ed. New York: Churchill Livingstone, 2008. p. 2560.

IWOOD. International Workshop on Definitions in Ontologies. 2014. *In: INTERNATIONAL CONFERENCE ON BIOMEDICAL ONTOLOGIES (ICBO)*, 5, 2014, **Proceedings...** Houston, TX, USA, October 7, 2014. Disponível em : <https://sites.google.com/site/definitionsinontologies/>. Access at: nov 2014.

KEIZER, Nicolette F.; ABU-HANNA, Ameen.; ZWETSLOOT-SCHONK, Johanna.H. Understanding terminological systems. I: Terminology and typology. **Methods of information in medicine**, v.39, n.1, p.16-21. Mar. 2000.

KÖHLER, Jacob *et al.* Quality control for terms and definitions in ontologies and taxonomies. **BioMed Central bioinformatics**, v.7, p. 212. Apr. 2006.

MEEHAN, Terrence, F. *et al.* Logical development of the cell ontology. **BioMed Central bioinformatics**, v.12, p.6, 2011.

MICHAEL, Joshua; MEJINO JUNIOR, Jose L.; ROSSE, Cornelius. The role of definitions in biomedical concept representation. *In: AMIA SYMPOSIUM, 2001. Proceedings...* [s.l.]; [s.n], 2001. p.463-7

NAJFELD, V. Conventional and molecular cytogenetic basis of hematology malignancies. *In: HOFFMAN, R. et al. (ed.). Hematology: basic principle and practice. 5th.ed. Churchill Livingstone: Elsevier, 2009. cap. 55, p.791-838.*

NATIONAL CANCER INSTITUTE (United States). **NCI dictionary of cancer terms: acute myeloid leukemia.** Available on: <http://www.cancer.gov/publications/dictionaries/cancer-terms?CdrID=46347>. Access at: 15 jul 2015.

OPEN BIOLOGICAL AND BIOMEDICAL ONTOLOGIES (OBO) Foundry. **Home page.** Availale on: <http://www.obofoundry.org/>. Access at: 20 maio de 2018.

ONTOBEE. **Leukemia.** 2014a. Availale on: [http://www.ontobee.org/browser/rdf.php?o=DOID&iri=http://purl.obolibrary.org/obo/DOID\\_1240](http://www.ontobee.org/browser/rdf.php?o=DOID&iri=http://purl.obolibrary.org/obo/DOID_1240). Access at: maio 2014.

ONTOBEE. **Welcome to Ontobee!** 2014b. Availale on: <http://www.ontobee.org/>. Access at: 5 maio 2014.

PETROVA, Alina; et al. Formalizing biomedical concepts from textual definitions. **Journal of biomedical semantics.** v.6, p.22. apr. 2015.

SCHRIML, Lynn M.; et al. Disease Ontology: a backbone for disease semantic integration. **Nucleic Acids Research.** v.40. 2012. Availale on: <http://nar.oxfordjournals.org/content/40/D1/D940.full.pdf?keytype=ref&ijkey=oOOVAj4zMzRpayq>. Access at: 21 set 2014.

SEPPÄLÄ, Selja; SCHREIBER, Yonatan; RUTTENBERG, Alan. Textual and logical definitions in ontologies. *In: INTERNATIONAL WORKSHOP ON DRUG INTERACTION KNOWLEDGE MANAGEMENT (DIKR 2014), THE SECOND INTERNATIONAL WORKSHOP ON DEFINITIONS IN ONTOLOGIES (IWOOD), 1., 2014, AND THE STARTING AN OBI-BASED BIOBANK ONTOLOGY WORKSHOP (OBIB 2014), CEUR Workshop Proceedings, v.1309, Houston, TX, USA, October 6-7, pp. 35-41.* Availale on: <https://seljaseppala.files.wordpress.com/2014/11/seppc3a4lc3a42014a.pdf>. Acesso em 21 maio 2015.

SEPPÄLÄ, Selja ; RUTTENBERG, Alan. Survey on Defining Practices in Ontologies : Report. in preparation of the International Workshop on Definitions in Ontologies (DO 2013). *In: INTERNATIONAL CONFERENCE ON BIOMEDICAL ONTOLOGY (ICBO 2013), 4, JULY 2013, Montreal, Quebec, Canada, 2013.* Availale on: [http://definitionsinontologies.weebly.com/uploads/1/7/6/9/17696103/do2013\\_surveyreport\\_seljapalaruttenberg.pdf](http://definitionsinontologies.weebly.com/uploads/1/7/6/9/17696103/do2013_surveyreport_seljapalaruttenberg.pdf). Access at: nov 2014.

SILVA, Grazielle C. da *et al.* Diagnóstico laboratorial das leucemias mieloides agudas. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, Rio de Janeiro, v. 42, n. 2, p. 77-84, abr. 2006 . Availale on:

[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1676-24442006000200004&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1676-24442006000200004&lng=en&nrm=iso). Access at: 26 ago. 2015.

SMITH, Barry. Introduction to the Logic of Definitions. *In: INTERNATIONAL WORKSHOP ON DEFINITIONS IN ONTOLOGIE*, DO 2013, July 7, Montreal, **Proceedings...** 2013. Availale on: [http://ceur-ws.org/Vol-1061/Paper5\\_DO2013.pdf](http://ceur-ws.org/Vol-1061/Paper5_DO2013.pdf). Access at: maio 2014.

SMITH, Barry *et al.* **Relations in biomedical ontologies**. *Genome biology*. v.6, n.5, p.R46.2005.

SOERGEL, Dagobert. **Knowledge organization systems: overview.2014**. Disponível em : <http://www.dsoergel.com/SoergelKOSOverview.pdf>. Access at: 14 ago 2014.

SPEAR, Andrew D. **Ontology for the twenty first century: An Introduction with Recommendations**. Institute for Formal Ontology and Medical Information Science (IFOMIS), Saarbrücken, Germany .University at Buffalo, Buffalo, New York, U.S.A. 2006. Disponível em : <http://ifomis.buffalo.edu/bfo/documents/manual.pdf>. Access at: 4 abril 2014.

SWARTZ, Norman. **Definitions, dictionaries, and meanings**. This revision: November 8, 2010. Disponível em : <http://www.sfu.ca/~swartz/definitions.htm>. **Acesso em meio 2014**.

THE GENE ONTOLOGY. **An introduction to the gene ontology**. Availale on: <http://www.geneontology.org/>. Access at: 20 nov 2013.

TSATSARONIS, George; et al. Learning Formal Definitions for Biomedical Concepts. *In: SRINIVAS, K; JUPP, S. (ed.). Proceedings of the 10th OWL: Experiences and Directions Workshop (OWLED 2013)*, May 2013. Availale on: <https://dill.inf.tu-dresden.de/web/LATPub509/en>. Access at: 21 maio 2015.

USCHOLD, Mike. **Building ontologies: Towards a Unified Methodology (1996)**. Availale on: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.39.9075>. Access at: 15 fev 2013.

VARDIMAN, James.W. *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. **Blood**. v.114, n.5, p.937-51. jul. 2009.

XIANG, Zoushuang; et al. Ontobee: A Linked Data Server and Browser for Ontology Terms. *In: INTERNATIONAL CONFERENCE ON BIOMEDICAL ONTOLOGIES (ICBO)*, 2nd. July 28-30, 2011, Buffalo, NY, USA. p.279-281. **Proceedings...** 2011. URL: <http://ceur-ws.org/Vol-833/paper48.pdf>.



Artigo submetido ao sistema de similaridade

Submitted: 11/10/2019 – Accepted: 10/02/2020 – Published: 12/05/2020

---