Information Polity 20 (2015) 119–134
DOI 10.3233/IP-150361
IOS Press 

# Visualizing *Aedes aegypti* infestation in urban areas: A case study on open government data mashups

Patricia Graziely Antunes de Mendonça[a], Cristiano Maciel[a,b,*] and José Viterbo[c]
[a]*Instituto de Educação, Programa de Pós-graduação em Educação, UFMT, Avenida Fernando Corrêa da Costa, Cuiabá, MT, Brasil*
[b]*Instituto de Computação, UFMT, Avenida Fernando Corrêa da Costa, Cuiabá, MT, Brasil*
[c]*Instituto de Computação, UFF, Rua Passo da Pátria, Bloco E, Niterói, RJ, Brasil*

**Abstract.** Publishing open government data and providing visualizations that favor the quick understanding and interpretation of the associated information is a task that requires systematization and the use of various technologies. Due to technical limitations (dependence on technological platforms, for example), this is not a standard practice among government system developers. In this work, we present a case study that involves the creation of a map to visualize information about the infestation of *Aedes aegypti* (dengue vector) in the municipality of Cuiabá, based on the selection and analysis of open data. This case study not only demonstrates that offering an easy way for presenting information on public health brings benefits to the community, but it also discusses a strategy for government agencies publishing open data and creating visualizations that combine information from various sources.

Keywords: Open government data, semantic web, linked data, data visualization

## 1. Introduction

Current governmental projects, aiming at the citizen's inclusion in the digital world, have stimulated the search for products, information and services made available by Information and Communication Technologies (ICT). The use of this set of technological resources intends to make citizens a key part of the reforms in the public sector [22].

The sheer publication of government data in open format is not enough to ensure citizen participation in monitoring government actions. For this to occur, it is necessary to use technologies for generating views that may be easily understood by the general public so that it acts as an agent of change, overseeing the government's actions. Even after finding the available information in the appropriate format, the user must be able to understand it and interpret it. Starting from this premise, easy to use tools for creating user-centered mashup services [16] are necessary.

*Corresponding author: Cristiano Maciel, Instituto de Computação, UFMT, Avenida Fernando Corrêa da Costa, 2367, CEP 78060-900, Cuiabá, MT, Brasil. Tel.: +55 65 39158383; Fax: +55 65 39158383; E-mail: cmaciel@ufmt.br.

According to [2], a mashup is a web application that combines content from many sources to produce something new with greatly enriched data visualization. In [3], it is stated that the information visualization area aims to enable countless techniques that allow effective visual representation of information (such as color, shape and size exploration). It also emphasizes that visualizations should be user-friendly in order to allow users to make new interpretations as well as widening comprehension.

In line with this reality, the Brazilian government is providing access to data for the general public, so that society can act as a supervising agent of public actions. For this purpose, the Brazilian government issued a law, known as Information Access Law [1], which became a huge mark as it stated that access to public information is a rule and secrecy an exception. This law also establishes some requirements to be fulfilled when publishing data, such as the availability in open and non-proprietary format, the use of standards that allow automated data access by machines and the timely provision of updated data, among others. Indeed, the success of publishing government data will occur when citizens become able to easily access, interpret and reuse such data.

By using semantic web technologies, organizations can make public information available online, providing meaning to each piece of data. This favors the merging of data and the creation of services to supply different types of visualization such as, for example, the use of tables, charts, maps, diagrams, mashups etc.

According to [23], Open Data *"refers to the possibility to make data, and particularly Public Administrations data, freely available (i.e., open) with no or limited restrictions on reuse"*. Publishing open data in order to favor the creation of visualizations that combine information from various sources is a task that demands systematization and the use of different technologies. The lack of a clear methodology to guide government system developers in this task has been inhibiting the conventional application of such techniques.

In related work, we found studies that focus on developing the concept of Open Government Data (OGD), government advances for opening data and the implementation of systems that enable the automated access to machine readable data.

Along these lines [19], proposed checking the transparency of public accounts on the basis of available data from the computerized public accounts audit system on the Citizen Portal of the City of Nova Mutum, Mato Grosso state, in the year 2011. One can notice that the systems developed by the government IT professionals did not help the citizens to understand the public accounts and most often hamper the presented information and views.

Also noteworthy is the approach by [20] who collected and selected information about corrupt politicians published on the JusBrasil[1] website. For this, a conversion was made of the relational databases into semantic databases in order to perform queries on the semantic web.

In [4], the authors proposed a strategy that consists of a sequence of steps for publishing data in an open format based on the linked data standard, using several associated technologies. Data published by the Public Security Department of Rio Grande do Sul State[2] served as input for this study referring to state crime statistics.

From these studies came the inspiration for proposing this case study: the use of a data set obtained in the Cuiabá Health Department (Secretaria Municipal de Saúde de Cuiabá) and in the State Health Department of Mato Grosso State (Secretaria Estadual de Saúde de Mato Grosso) about *Aedes Aegypti*

[1] Available in: http://www.jusbrasil.com.br/bem-vindo?ref=top.
[2] Available in: http://www.ssp.rs.gov.br/?model=conteudo&menu=196&pg=1.

(dengue fever vector) infestation in the city of Cuiabá. This study aims also at testing the strategy proposed by [4] which consists of a sequence of steps for publishing data in an open format based on the linked data standard, duly analyzing and adapting it to the specific purpose at hand. The strategy is extended to include a new step after the open data is published: the creation of visualizations that integrate data from different sources, especially using mashups. This work is organized as follows. After the introduction, we present a brief theoretical reference. In Section 3, we present the case study and in Section 4 we draw our final considerations.

### *1.1. Motivation*

According to information from the Brazilian Ministry of Health [24], Dengue fever is an infectious disease caused by a virus. According to [25] "Dengue fever is transmitted by the bite of an infected mosquito [...]". The *Aedes aegypti* mosquito is the disease vector and originates from Africa arriving in Brazil during the colonial period. This mosquito appears mainly in tropical and sub-tropical parts of the planet, such as in Brazil where the climatic conditions allowed the insect to adapt well into all regions of the country. *"In the Americas, the disease has spread through cyclical outbreaks occurring every 3–5 years"* [24].

In this way, the disease now exists all year round in Brazil with epidemics occurring in rainy periods. [24] explains that the mosquitos hide in places where water has accumulated and lays its eggs there. These places are known as "hatcheries". And if no water is present, they can remain in the hatchery for up to one year. In these periods, whenever there is just a single drop of water, the eggs can hatch and thus the mosquito procreates. The largest incidence of the disease is in rainy periods when there is greater vector reproduction.

Dengue fever presents symptoms such as high fever, headache, pain behind the eyes, backache, red blotches over the body and may cause hemorrhaging which can lead to death [24].

Thus, with the aim of reducing these annual outbreaks of Dengue, state and federal governments have performed awareness-raising campaigns to encourage prevention of mosquito proliferation.

However, these efforts aiming to eliminate the Dengue vector have not been very effective. In fact, it can be seen from the Epidemiological Bulletins published by the Brazilian Health Ministry that the lack of public participation in prevention and caring is very high leading to an increase in the disease incidence rates in all regions of the country as shown in [24].

In this country, it is very easy to find someone who has had the disease as anyone who is bitten by the vector can contract it. And this sad reality of the high expansion rate can be seen daily in newspapers. Thus, we believe that providing access to information on the increase of Dengue in the city of Cuiabá can aid the combat and prevention efforts in a city that suffered a very large outbreak in 2009 which caused several deaths.

The survey on Aedes aegypti mosquito infestation is done using the Rapid Infestation Rate Survey (LIRAa) which is published every two months.

Following up from this, this topic is justified by the importance of allowing public access to the information obtained from the efforts made by the Cuiabá Department of Health which runs the Dengue Combat Program and which can provide open government data on the disease and vector infestations, the LIRAa on the *Aedes aegypti* mosquito, in the city.

In the state of Mato Grosso, there are still no reports on the use of Open Government Data and in the city of Cuiabá when data is available they are restricted to certain formats that do not allow for manipulation in order to generate information, such as Comma Separated Values (CSV) files that do not meet the principles of Open Government Data. In other cases, the data are found in *Excel* spreadsheets and even then they are mostly provided in printed form.

This way, we believe that the possibility of visualizing data graphically, in an easy to understand format, can help civil society to have a global view of the Dengue situation in the city of Cuiabá. This can favor a change of attitude by the public encouraging people to act in a more collaborative way in order to eliminate the hatcheries.

However, the publication of data in an open format alone is not enough to ensure the public, civil participation. But, it could become a powerful tool to support permanent combat of the disease by the Health Department.

### *1.2. Methodology*

This study was initiated with a bibliographical survey by means of reading and analysis of books, journals and technical guides in the areas of Open Government Data, Linked Data, ontologies, web semantics, information visualization and mashups.

Further, an explanatory survey was carried out using a case study and qualitative observations based on documentary analysis on Dengue data in the city of Cuiabá, Mato Grosso. For [29], explanatory cases are suitable for doing causal studies, mainly to test theories.

Besides this, this study is of an experimental nature since it aimed to test the use of strategies for publishing Open Data in the Linked Data standard proposed by [4], seeking a "short-cut" for its application to allowing publication of the use of Open Government Data. For this, it was necessary to study a variety of tools used from the data treatment phase up to the mashup production phase seeking to illustrate and discuss the advantages of publishing open data.

## 2. Fundamental concepts

The case study described in this work involves several concepts, such as open government data, linked data and mashups, which are explained in this section.

### *2.1. Open government data*

Open government data are data produced by the government and made available to people not only for reading and informational purposes but also to be re-used in new projects, sites and applications [5]. The simple publishing of data in proprietary format, for example, does not ensure its applicability. Even if the publication intends to promote public transparency, some rules are necessary. For the [6], open government data must fulfill eight basic principles, which assume that they must be: complete, primary, updated, accessible, machine understandable, non-discriminatory, non-proprietary and license free. Attentive to these principles, [7] observed that the web has changed and that by using linked data, it will be extended through common data which will connect data from different sources.

Open data publishing using web semantic technologies facilitates handling since they can be easily understood by machines which further facilitates automatic information visualization and the construction of new knowledge [8] highlights that web semantics allow computers to have qualitative information using ontologies to point out the vocabulary used in semantic applications. An ontology is an explicit knowledge model, a formal way to declare people's names, places and things along with their descriptions properties and relationships [8]. In linked data simple ontologies are used to supply meaning to available data [4].

### *2.2. Linked data*

The World Wide Web Consortium (W3C) recommends the use of linked data for publishing open government data [9]. This uses semantic web technologies to publish structured data on the web and to interconnect data sources, linking the data "to the outside", to other data and other documents, with the aim of extending the web with common data, integrating data from different sources.

Such connections are useful in capturing non-hierarchical relationships since they are bidirectional, enabling consultations that, from different perspectives, allow flexible and complete research that extends throughout many documents and sites [10].

[7] gives four rules for linked data publishing:

1. Use URI's as the names of things;
2. Use HTTP URI's so that people can search for names;
3. Supply useful information when someone searches for a URI using standards (Resource Description Framework (RDF), SPARQL Protocol and RDF Query Language (SPARQL)); and
4. Include links to other URI's, so that more information can be discovered.

If those rules are obeyed, web content will have a strongly organized structure based on consultation patterns and data recovery (RDF, Ontology Web Language (OWL), SPARQL), that provide the basis for the creation and application of more intelligent tools.

Through these guidelines, when using the linked data standard the necessary requirements will be met for consultation with open government data as well as allowing the production and maintenance of collaborative information [11]. However, methodologies for the use of open government data with standards have begun to emerge, as discussed by the above authors. For them, "Methods for producing and publishing open government data are required and need further research." However, [28] state that "Linked data lets us realize open data's benefits with less effort from users. RDF supports several data models, so the data can be retrieved not only as RDF, but also as JavaScript Object Notation (JSON) or CSV, for example".

One of these is the strategy proposed by [4] to publish open government data in the linked data standard. The authors present a set of steps that compose the process, which are represented in Fig. 1.

In that study [4], the author proposed a series of steps with the purpose of publishing open government data on the Linked Data standard. The model proposes the following phases:

1. Obtaining the data in CSV format (Department of Public Security of Rio Grande do Sul). These were data about the statistics of police occurrences that were already open in eXtensible Markup Language (XML) and CSV format, filtered by county and by month, within the year and describing the occurrence type;
2. Migrating to a relational database. The database was normalized and logical, conceptual and physical models were developed using the DBDesigner tool, in order to separate the terms contained in the data;
3. Development of ontologies utilizing the Protégé, a free, open-source ontology editor.[3] After searching for an ontology focused on the Brazilian penal code in order to reuse it and not finding something that met this need, a simplified ontology was developed [4], "with low semantic value" with the objective of publishing open government data;

[3]Available in: http://protege.stanford.edu/.

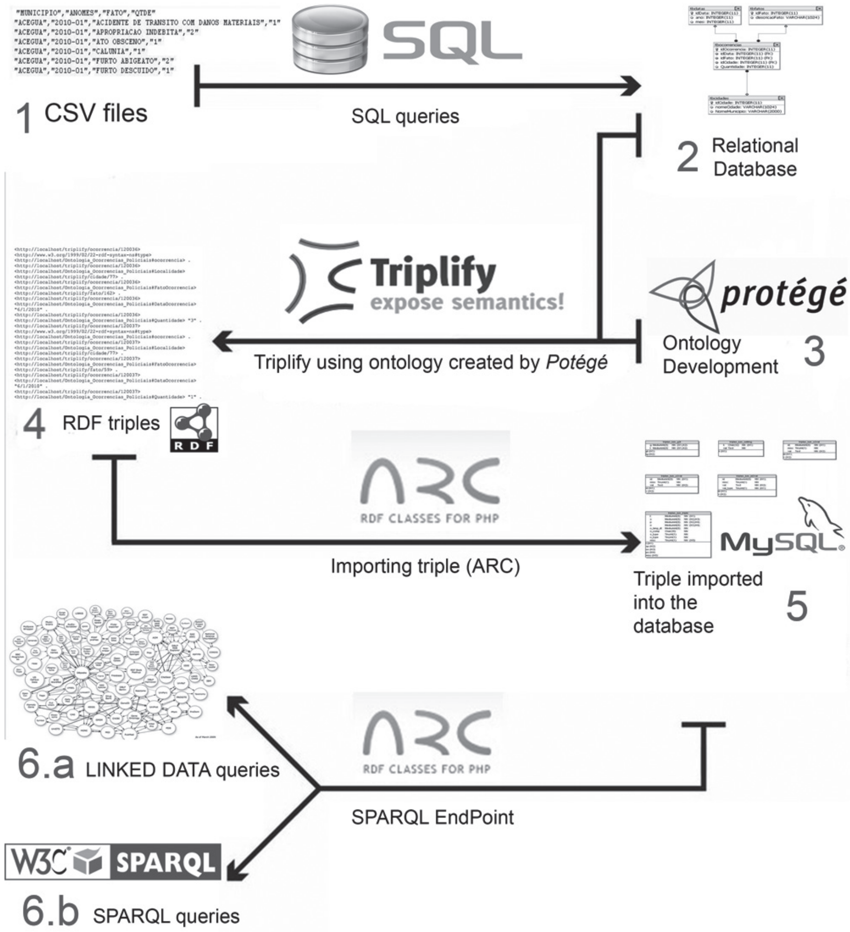


Fig. 1. Summarization of the tasks proposed by [4].

4. Generation of triple RDF. Aiming to provide meaning to the data from the relational database, the information about these data and its metadata (created ontology) were connected, semi-automatically, through the *Triplify* software. This tool has allowed access to a relational database to generate the triple RDF;
5. Importing the triples for ARC2 server, a flexible RDF system for semantic web and PHP practitioners. This tool is free, open-source, easy to use, and runs in most web server environments (it is PHP 5.3 E_STRICT-compliant)[4] Utilizing the converting function to the database, the triple were read and uploaded to the server, which already generated the tables, where the RDF file data were stored. These have been made available for access through the SPARQL language, utilizing the functions of the triples repository and the SPARQL EndPoint contained in the ARC2 server;
6. Providing the triple access in the same ARC2 server. After providing access to the triples, in the last development phase of his work [4] suggested two examples of data access which could be realized through Linked Data (point 6a in Fig. 1) in RDF and would be realized through query using the

[4] Available in: http://semanticweb.org.

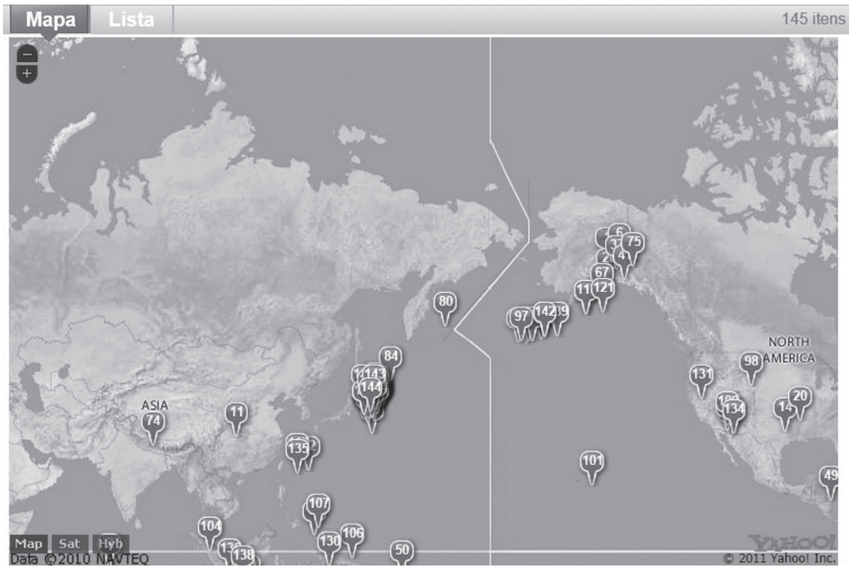


Fig. 2. Visualization using Yahoo Pipes.

SPARQL language after the publication of the information on a host with a fixed internet address. And in phase 6b in Fig. 1, he suggested that the triple were accessed for consultation on SPARQL EndPoint.

Such steps are applied in the case study within this article, since they serve as a basis for the study's objective, discussing this model and adding value to the visualization of open format data.

### *2.3. Open data and mashups*

From procurement of open government data, it becomes easier to create applications that use this information, "since the data sets are created and updated by the organizations responsible for the information" [11]. The presentation of information in a graphical format, in addition to maximizing the use of automatic data analysis techniques, makes it palatable to consumers of government services.

Several studies in the literature are using mashups for data visualization, but the focuses of these works are not on data type and there is low use of open data format.

For [12], mashups are software applications that mash up Application Programming Interfaces (API) and data sources in an integrated interface to create the desired visualization.

In Fig. 2, an example is presented of a mashup developed in Yahoo Pipes[5] tool that graphically shows regions of the world where earthquakes have occurred during the last seven days.

The concept of mashup for the web is not considered new because since the emergence of the web many implementations have been created by mixing codes in HyperText Markup Language (HTML) with information from multiple sources creating new formats, new documents and applications with increased usability.

An example of work that allows visualization of public data is in the "Para onde foi o meu dinheiro?" (Where did my money go?) website, an application available by the Nossa São Paulo (Our São Paulo)

[5] Available in: http://pipes.yahoo.com/pipes/.

network along with W3C. The application aims to help the public monitor the execution of city, state and federal budgets.

Thus, it is confirmed that mashups are created primarily for more enriched data providing individuals, organizations and governments with the generation of new meanings.

## 3. Case study

This research develops a case study from the selection and analysis of data on the infestation of *Aedes aegypti* (dengue fever vector) in Cuiabá so as to display this information graphically on a map using open government data. The strategy used to build this visualization was similar to that proposed by [4]. However, there are differences in the way data was collected and processed and a new step was proposed.

### *3.1. Data collection for Linked data publishing*

The first step was the collection of data from two sources: the Cuiabá Health Department and the Mato Grosso's State Health Department.

During this process, we noticed that there is virtually no voluntary data disclosure on health in the municipality of Cuiabá, except for that reported by newscasts. As occurs throughout Brazil, much of the information and knowledge is still documented on paper. The few digital repositories that exist have different formats that often do not allow automatic processing, such as data available in PDF. For this organ, the data request was made by letter and the data was provided in the form of some printed worksheets and other in excel.

The delivery of the remaining data, supplied by the second source, was taken through the Data Warehouse Web[6] application which collects information on disease notification and any damage to physical, mental and social health of individuals caused by harmful circumstances. This data, which is stored in a database of the Brazilian Ministry of Health, is available in CSV format which theoretically would facilitate the development of this work.

Throughout Mato Grosso there are no reports of the use of open government data, although there are some available for the population, which are restricted to certain file formats that do not allow automated manipulation. Most of the time, data is available only in printed format which derails the systematization of such information. Besides, there are no known ontologies for presenting such data, making it impossible to re-use them or to create new extensions for them.

Despite this outdated scenario, there is a federal government initiative making efforts to develop repositories such as Vocabularies and Ontologies for Electronic Government[7] (e-VoG) which aims to facilitate the exchange of information with semantic agreement, favoring the crossing of data from various sources and still enabling the reuse of existing ontologies.

### *3.2. Creation of a relational database*

In the next step, we performed a study of the data dictionary of the Data Warehouse Web system, aiming at creating a relational database containing every concept defined in the dictionary.

[6] Available in: http://www.saude.mt.gov.br/sistemas.
[7] Available in: http://vocab.e.gov.br/2011/03/vcge#esquema.

Furthermore, the data obtained in printed form were manually included in another relational database.

The data preparation phase of preparation was important to identify if the data reported by the Cuiabá Health Department was more comprehensive. During the inclusion of such data in the relational database, it has been confirmed that they met most of principles outlined in [6], being comprehensive, primary, up-to-date, non-discriminatory and license free. Therefore, they were ready to be converted into open government data, thereby also becoming accessible, understandable by machines and non-proprietary.

On the other hand, the data from the Data Warehouse Web was incomplete, not attending the basic principles of [6]. After importing such data to a relational database, it would be necessary to handle the information and update the files in order to utilize them in all phases of the study. So, due to the fact that such data has poor quality, is not comprehensive and is not primary, it was discarded because it could not become open government data and be presented as having been collected from the source [6].

### *3.3. Ontology creation*

Prior to the creation of ontologies, a search was performed for ontologies about Dengue fever and the Levantamento do Índice Rápido de Infestação do *Aedes aegypti* (*Aedes aegypti* Fast Survey Index – LIRAa) in cities where LIRAa was identified per neighborhood. However, no available ontologies were found, with the adhering characteristics to the LIRAa to be extended and reused. So, the next step was to create a LIRAa ontology.

As proposed in [4], we started the development of a simplified ontology in order to perform the necessary steps to fulfil the purpose of this paper.

Initially, the ontology was manually developed, following the Methodology for Building Ontologies developed by the group of Requirements Engineering from the PUC-Rio. This methodology is described in [21], which presents a simplified version of this method based on Language Extended Lexicon (LEL), where a list of related terms was created. However, since the conclusion of the manual mapping is hard and difficult [13], and out of the focus of this study, from this point we followed in a semi-automatic way, using the Protégé tool.

In this step, data were entered in a Protégé tool (version 3.4.7), with the purpose of creating an ontology called OntoLIRAa (LIRAa ontology) containing the Estrato and LIRAa classes (Fig. 3), which refer to the tb_estrato and tb_liraa tables of the relational database. This implementation confirms what was shown by [13]: a table is mapped to a class and the columns are mapped as datatype properties.

Properties that represent the attributes of the LIRAa database data were added to the construction of classes. At this time, it was found that, for this data set, it would not require the creation of forums for the transformation of data in the relational database into RDF triples. Thus, for this situation, it was enough to create Datatype and Object Properties without the need for inclusion of instances which refer to the database triples.

### *3.4. Data triplification*

In the fourth step, we used the Triplify software that enables the creation of mapping of the relational data by means of SQL (Structured Query Language) sentences to identify the properties of the database, transforming them into RDF triples, JSON and Linked Data ready for use in web applications [14].

Upon triplification the database is automatically converted and receives metadata reported in the step prior to the creation of ontologies.

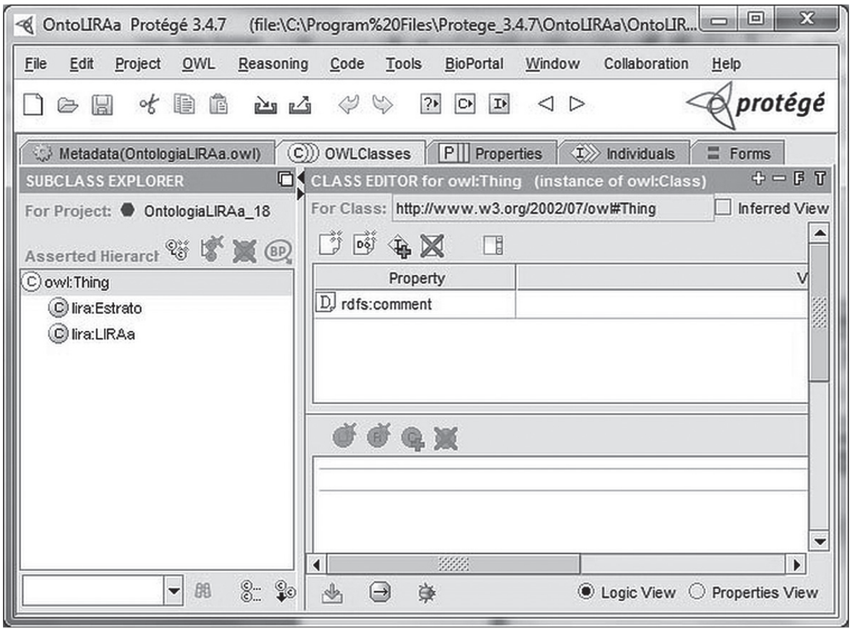


Fig. 3. Classes created in Protégé tool.

### 3.5. Availability of open data

After triplification, fulfilling the fifth step, the RDF triples are imported into the database called Trebles, created in MySQL. Then, in the sixth step, we used the triple store and SPARQL EndPoint available on the ARC2 server, thus providing the triples generated for consultation.

According to [15], the triple repository is used to store and query RDF data, with the latter only becoming a viable activity as a consequence of the ARC2 service available on the server, the SPARQL EndPoint, which implements the SPARQL protocol.

Automatically, the ARC2 server creates the tables (Fig. 4) from the database and populates it.

Using the SPARQL EndPoint, there was a query to the database containing the RDF triples using the SPARQL language. A machine readable result was obtained that can be made available in JSON, RDF and other formats.

### 3.6. Open data visualization on a map

After converting the relational database into RDF triples and subsequent publication by the ARC2 server, the final step of this work was achieved visualizing data in open format which can now be used in several applications and even merged with other data as well.

Before the graphical representation of data, the performance of both procedures was needed: the execution of SPARQL queries for data selection, and compliance with the requirements of the database created automatically during import of triples. When performing the query in SPARQL EndPoint, you can select the format of the result to be obtained. The application offers output options in XML, JSON, Plain, Serialized PHP, Turtle, RDF/XML, Query Structure, HTML Table and TSV formats.

In the present study, the format used was JSON and the query response was applied to one of the models available for viewing information on the API Exhibit.

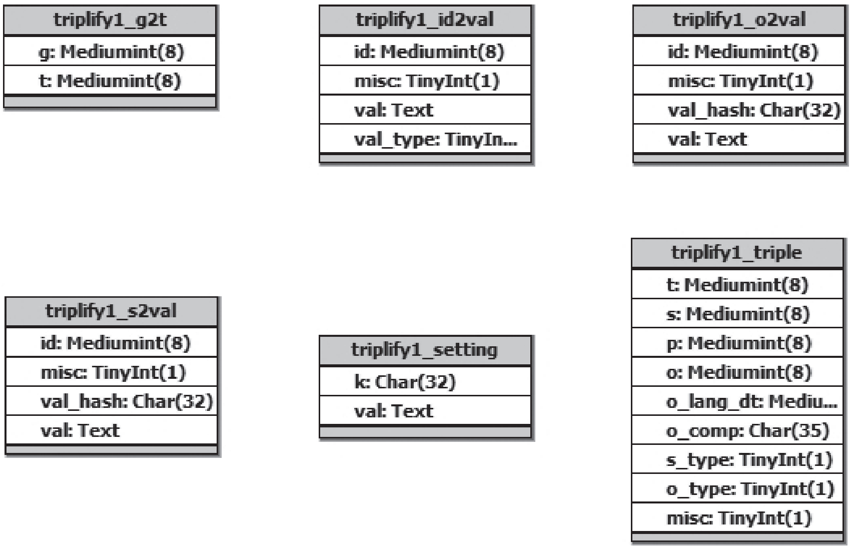


Fig. 4. Tables created by the BD converter of ARC2 server.

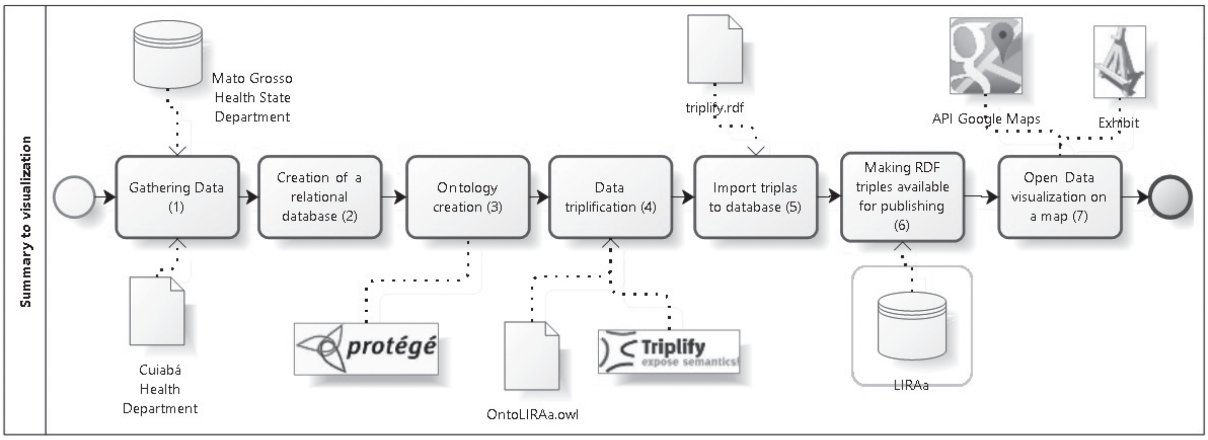


Fig. 5. Summary of the *Aedes aegypti* infestation data visualiation steps.

After performing all of the steps presented in the previous stages, a graphic representation of these data was included, generating information on a city map.

Figure 5 synthesizes all steps performed in this study, including the visualization of information.

The API Exhibit[8] is part of a large open-source project of the semantic web, called SIMILE Widgets (Semantic Interoperability of Metadata and Information in unLike Environments Widgets), developed at MIT – Massachusetts Institute of Technology. The Simile project has focused on developing free, robust, open source tools and which enable access, management, visualization and reuse of digital assets. In the scope of this project, different API's were produced and Exhibit is one of them.

When defining how to display data in this study, we selected a U.S. map view that originally contained

[8]Avaliable in: http://www.simile-widgets.org/exhibit/.

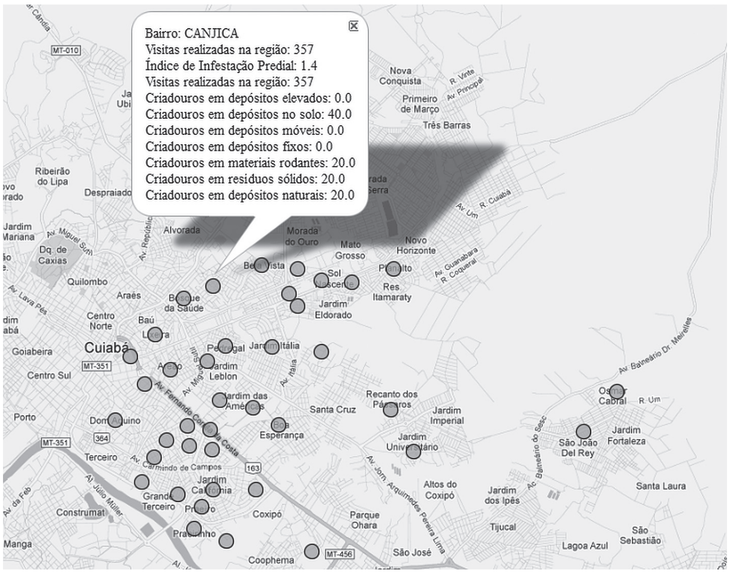


Fig. 6. Visualization of *Aedes aegypti* infestation open data.

the cities identified and associated with the number of inhabitants. This model was reused to present data of the infestation of *Aedes aegypti* in Cuiabá.

The reuse of the API is quite simple: you can use a text editor and very little knowledge of HTML. After downloading the source code and file with the available data and then saving them in a directory, simply populate the data file with the outcome of a consultation on SPARQL Endpoint obtained in JSON format. The preview is displayed in a browser (Fig. 6).

### *3.7. Strategy analysis*

In this section, we analyze the steps proposed for carrying out this study, which primarily aims at testing the strategy for the publication of open data in the linked data standard, followed by its visual representation using mashups, proposed in [4].

In Table 1, the steps of the study are presented in a summary, bearing in mind the goals, the inputs and the results obtained for each step.

However, it is necessary to explain some of the peculiarities that occurred during the development:

- Data capture. Differently from the data collected by [4], which was in CSV format, the data in this study was distinct in that it was obtained in printed, Excel and CVS formats. Furthermore, the CVS files could not be used in the visualization because they were incomplete and broke the principles of linked data.
- Creation of a relational base. Two databases were created where one was for the data from the Mato Grosso State Department of Health and the other was for the data from the Zoonosis Control Center. For the reasons already stated, only one database was used.
- Creation of ontologies with the Protégé tool. In this step, with a view to experimenting alternatives to the [4] proposal, two files were created called OntoLIRAa.owl, whereby one contained only the classes, datatypes and object properties and the other also included the 234 instances for Estrato and 120 for LIRAa.

Table 1
The steps necessary for creating the visualization of the Aedes aegypti infestation data

| Step/task | Goal | Input | Output/Result |
|---|---|---|---|
| 01 | Data collection for performing the study. | CSV files from the State Health Department; and electronic and printed spreadsheets from the Cuiabá Health Department. | Inserting data into the *MySQL* database. |
| 02 | Creation of a relational base. | Non-normalized data in the database. Data from the State Health Department are inconsistent. | Normalized data in the database from the Cuiabá Health Department. |
| 03 | Creation of ontologies using the *Protégé* tool. | Relational database with data from LIRAa – "Fast Survey *Aedes aegypti* Index" | OntoLIRAa.owl (LIRAa ontology) |
| 04 | Creation of RDF triples using the *Triplify* tool. | OntoLIRAa (LIRAa ontology) – file in the OWL format. | triplify.rdf – file containing the RDF triples. |
| 05 | Importation of the triples into the *MySQL* database. | triplify.rdf file | LIRAa, database in graphs |
| 06 | Provision of the RDF triples for publication with the ARC2 server. | LIRAa, database in graphs | *SPARQL EndPoint* consultations. |
| 07 | Data visualizations in mashups. | *SPARQL EndPoint* consultations | Visulaization of the data in a Google map using the API Exhibit. |

– Creation of the RDF triples using the Triplify tool. With the aim of testing the result of the ontologies previously created, we excluded two conversions of databases into RDF triples using the two OntoLIRAa.owl files created in the previous step. Thus, it was proven that the tool behavior was not altered.

The steps corresponding to importing the RDF triples from the MySQL database and provision of the RDF triples for publication were fully completed as in [4]'s proposal. Also, after testing the strategy, there was a display of the data by means of data visualizations using mashups and reusing the *API Exhibit.*

After the data collected was made available in RDF, it was only necessary to manipulate it to generate the desired visualization. However, the greatest difficulties were associated with the data and information collection. This was because the data originating from the Data Warehouse Web (of the State Health Department) were quite incomplete containing incorrect information in relation to the location of Aedes aegypti infestations (the city neighborhoods were mostly missing from the persisted tuples), which caused mistaken visualizations of the data on the map if these hadn't already been discarded in the relational database creation phase.

Furthermore, information was missing that could clarify the concepts inserted in the data collected so that afterwards, when the data were made available in Linked Data you could also document them in accordance with the orientation of [26] in terms of the need to follow best practice for publication.

With the aim of clarifying these practices [26], presents a checklist with eight items to be carried out so that the publication presents self-descriptive data which help client applications to integrate data from different sources. These are:

1. Does your data set link to other data sets?
2. Do you provide provenance metadata?
3. Do you provide licensing metadata?
4. Do you use terms from widely deployed vocabularies?
5. Are the URIs of proprietary vocabulary terms dereferenceable?

6. Do you map proprietary vocabulary terms to other vocabularies?
7. Do you provide data set-level metadata?
8. Do you refer to additional access methods?

Thus [26], believes that by answering these questions they would establish aptness for publication of quality Linked Data that favors reuse.

In conclusion of this analysis, it was also possible to identify that [4]'s study is in line with [26]'s directives in relation to the basic considerations of how to use URIs as the name for Things, describing Things with RDF and also following the "recipe" for serving Linked Data of Relational Databases. However, with the aim of improving this strategy, it was necessary to follow the orientations so that the data were self-descriptive about their restrictions so that the provided easier understanding for the Linked Data applications.

## 4. Final considerations

Open data is a recurrent topic especially in the realms of government. Therefore, there are some groups of professionals dedicated to the job of spreading this technology and applying it in such a way as to be useful to the public.

However, opening up data still requires additional efforts beyond the transformation of common data into open data as occurred with the data received from the Mato Grosso State Department of Health which were in CSV format, but which were not ready to be converted into open data. In other cases, the poor quality of data is an obstacle since its use breaks the principle of Open Government Data.

Among the lessons learned from Data.gov.uk [27], the definition of rules for data reuse is crucial since they need to have an open license in order to be considered open data. But, data managers do not have the authority to determine the ways of releasing data nor the policies, making this the greatest obstacle to OGD.

Besides the problem presented in [27], it was possible, during the research, to identify a strong resistance by governments in relation to the publication of open data since they are concerned with their use by the public and by other parties that could access primary data (collected from the source) and could manipulate them according to their own interests. This way, the governments would lose control of how these data are made available to the press. Should there be any interest from the local government, the OntoLIRAa could be reused every two months for the publication of data on *Aedes ageypti* infestations.

In this study, the greatest lesson learned is that the documentation of available data is of the utmost importance because, in order to create useful visualizations for the public, the data description needs to be clear in a document so that the visualizations present the real meaning of the data made available.

In this case study, we looked, in particular, at how the possibility of visualizing data in an easily understandable format can help civil society to have an overview of the situation of Dengue fever in Cuiabá. However, the systematization and publication of open government data on the infestation of *Aedes aegypti* still requires efforts to be fully achieved given the initial difficulty of data collection which had not even been gathered in a database.

As stated in [17], much of the information and knowledge remains on paper, except for a small portion already scanned, consisting of repositories in different formats that often do not communicate with each other. However, with the government's concern, and greater care to provide data that are ready to be reused, delays in reaching open data will soon be reduced. If the collected data was available in RDF format and there was an ontology to be reused, it would only be necessary to implement the steps defined from step 5 onwards, as shown in Table 1.

In such circumstances, there is a tendency for the federal government to alleviate this shortcoming through e-VoG, with respect to data from the Federal sphere, enabling broad searching for information from the public sector with the advantage of being simpler.

Regarding the objectives of this paper, we can say that they have been achieved and new research opportunities have been highlighted. It was possible to verify that the strategy proposed by [4] may be reused in a different context, given the adaptation of some steps. Examples of such adaptation occurred in creating the relational database step, where there was the need for manual input of data, the difference in procedures on creating a simplified ontology without inclusion of instances (which refer to the tuples in the database) because the ontology was only required for the transformation of relational data into RDF triples. In addition, the main adaptation that differentiates this work from the original strategy consists of the final data presentation. Information is presented through a visualization map of the city of Cuiabá, Mato Grosso, where data on *Aedes aegypti* infestation is presented.

Also, it is necessary to examine some technologies carefully, such as SPARQL EndPoint tool, allowing you to do more sophisticated queries, such as selecting all the necessary content in order to populate the map of the city of Cuiabá, Mato Grosso, with all the information that can be found in the base with the semantic web, enabling the creation of further mashups.

The deepening of the study in other API's for viewing open government data can go further, since there is interest in the publication of these data in graphic form to be performed on social networks and even the institutional site of the Cuiabá Department of Health.

In this sense, there was great progress towards transparency [18], but there is still a lack of access to data that can be manipulated by machines. In the context of states and municipalities, the primary objective is to achieve transparency of actions. Many data are available in PDF files. Although the Federal Government has been striving to produce data for open disclosure of its procedures, methodologies and repositories, there is much to be done. Even though there is a law which deals with access to data released by the government, most public institutions do not provide data in an open format, which makes it time consuming for developers of solutions and thus creates a barrier to popular support.

It is believed that experiments such as this one, which presents and discusses methodologies for producing, processing and using open government data, need to be socialized and deserve consideration in the scientific community and in the domestic market in order to achieve effective progress in these areas. To this end specific methods, techniques, processes and tools must emerge from government spaces in which such use has been sought, fostering partnerships with universities, companies and institutions that allow for the proposal of new solutions and investigate the use of such data by the general public.

## Acknowledgments

We would like to thank the programs in Post-graduation in Education (PPGE/UFMT), Database specialization (IC/UFMT), Interactive Virtual Environment Laboratory (LAVI), Secretaria de Estado de Educação de Mato Grosso (SEDUC) and Uniselva Foundation for all the support to develop and publish this research.

## References

[1] BRASIL. Lei n° 12.527, de 18 de novembro de 2011. Regula o acesso a informações. *Diário Oficial da República*

*Federativa do Brasil.* Poder Executivo, Brasília, DF. Brasília, 18 nov. 2011 (in portuguese). Avaliable in: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 27/jun/2013.
[2] Meza, J. and Zhu, Q. 2008. Mix, Match, Rediscover: A Mashup Experiment of Knowledge Organization in an Enterprise Environment. In: International Journal of Knowledge Management (IJKM) no. 4 (1):65-76.
[3] Da Silva, I. C. S. and Freitas, C. MDS. 2006. Avaliação de ferramentas de busca na web baseadas em visualização de informações. In: VII Simpósio sobre Fatores Humanos em Sistemas Computacionais. Natal, Rio Grande do Norte, Brasil. ACM, 2006, p.178-185 (in portuguese).
[4] Baldus, L.; Maciel, C. 2011. Publicação de dados governamentais abertos no padrão linked data. Oficina. In: IV Congresso Internacional de Software Livre e Governo Eletrônico, Brasília, DF (in portuguese).
[5] Alonso, J. et al. 2012. Improving Access to Government through Better Use of the Web. W3C Interest Group, 12 May 2009. Avaible in: http://www.w3.org/TR/egov-improving/.
[6] Opengovdata.org website. 8 Principles of Open Government Data. 2007. Avaible in: http://www.opengovdata.org/home/8principles.
[7] Berners-Lee, T. 2006. Design issues: linked data. Avaible in: http://www.w3.org/DesignIssues/LinkedData.html.
[8] Siegel, D. 2010. Pull: o futuro da internet e o impacto da web semântica em seus negócios. Rio de Janeiro, RJ: Elsevier.
[9] Bizer, C., Heath, T., Ayers, D. and Raimond, Y. 2007. Interlinking Open Data on the Web. Demonstrations Track. In: 4th European Semantic Web Conference. Innsbruck, Austria, 2007. Avaible in: http://www.eswc2007.org/pdf/demo-pdf/LinkingOpenData.pdf.
[10] Graves, M. 2003. Projeto de banco de dados com XML. São Paulo, SP: Pearson.
[11] Maciel, C.; Viterbo, J.; Breitman, K. 2012. Transparência pública de dados abertos governamentais. In: RESENDE, A. (Org.). Governo brasileiro no futuro: Sugestões e desafios para o Estado (2012-2022). São Paulo: Cubzac, 2012, 47-56 (*in Portuguese*).
[12] Zang, N., Rosson, M. B., and Nasser, V. 2008. Mashups: Who? What? Why? In: Conference on Human Factors in Computing Systems. Florença, Itália: ACM. 3171-3176.
[13] Astrova, I. 2009. Rules for mapping SQL relational databases to OWL. In: MIGUEL-ANGEL Sicilia; MILTIADIS, D. Lytras (ED) Metadata and Semantics. New York, USA: Springer. 415-424.
[14] Auer, S. et al. 2009. Triplify: Light-weight linked data publication from relational databases. In: Proceedings of the 18th international conference on World wide web. Madrid, Espanha. ACM. 621-630.
[15] Breitman, K., Casanova, M.A. and Truszkowski, W., 2007. Semantic web: Concepts, technologies and applications. Springer.
[16] Väänänen-Vainio-Mattila, K., and Wäljas, M. 2011. Towards user-centered mashups: Exploring user needs for composite web services. In: Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems. Vancouver, Canadá. ACM. 1327-1332.
[17] Parastidis, S. 2011. Uma plataforma para estudo que sabemos: Criando uma infraestrutura de pesquisa orientada ao conhecimento. In: O quarto paradigma: Descobertas científicas na era da eScience. São Paulo, SP: Oficina de textos.
[18] Breitman, K., Salas, P., Casanova, M. A., Viterbo, J., Chaves, M. (2012). Open government data in Brazil. IEEE Intelligent Systems, 27(3), 45-49.
[19] Oliveira, L. K. B.; Maciel, C. Transparência das contas públicas e o controle social: Um estudo de caso com o Portal do Cidadão. Especialização em Controladoria e Finanças Empresariais. UNINOVA. Nova Mutum – MT, 2011.
[20] Carneiro, S. de O. Um método para recuperação da informação na web semântica utilizando a triplificação de dados. Trabalho de conclusão de curso. Engenharia de Sistemas Web, UFMT. Cuiabá – MT, 2012.
[21] Breitman, K. Web semântica: A internet do futuro. Rio de Janeiro (*in Portuguese*). LTC. 2005.
[22] Harrison, Teresa M., Theresa A. Pardo, and Meghan Cook. Creating Open Government Ecosystems: A Research and Development Agenda. Future Internet, 4 (4): 900-928. 2012.
[23] Corradi, A., Foschini, L., and Ianniello, R. 2014. Linked data for Open Government: The case of Bologna. *Computers and Communication (ISCC), 2014 IEEE Symposium on, vol., no., pp.1,7, 23-26 June 2014*. doi: 10.1109/ISCC.2014.6912473.
[24] Ministério da Saúde. Portal da Saúde. Avaible in: http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/ secretarias/svs/dengue.
[25] World Health Organization. Health topics, Dengue. Avaible in: http://www.who.int/topics/dengue/en/.
[26] Heath, Tom, and Christian Bizer. "Linked data: Evolving the web into a global data space." Synthesis lectures on the semantic web: Theory and technology 1.1 (2011): 1-136.
[27] Shadbolt, Nigel, O'Hara, Kieron, Berners-Lee, Tim, Gibbins, Nicholas, Glaser, Hugh, Hall, Wendy and schraefel, m.c. (2012) "Linked open government data: Lessons from Data.gov.uk." IEEE Intelligent Systems, 27, (3), Spring Issue, 16-24. (doi:10.1109/MIS.2012.23).
[28] Shadbolt, Nigel, and Kieron O'Hara. "Linked data in government." IEEE Internet Computing 4 (2013): 72-77.
[29] Dubé, Line, and Guy Paré. "Rigor in information systems positivist case research: Current practices, trends, and recommendations". MIS Quarterly (2003): 597-636.