

Publishing and Interlinking the Global Health Observatory Dataset

Towards increasing transparency in Global Health

Editor(s): Pascal Hitzler, Kno.e.sis Center, Wright State University, Dayton, OH, USA; Krzysztof Janowicz, University of California, Santa Barbara, USA

Solicited review(s): Natasha Noy, Stanford Center for Biomedical Informatics Research, USA; Willem van Hage, VU Amsterdam, The Netherlands; Giovanni Tummarello, DERI Galway, Ireland; one anonymous reviewer

Amrapali Zaveri^{a,*}, Jens Lehmann^a, Sören Auer^a, Mofeed M. Hassan^a, Mohamed A. Sherif^a and Michael Martin^a

^a *Universität Leipzig, Institut für Informatik, AKSW, Postfach 100920, D-04009 Leipzig, Germany*
E-mail: {lastname}@informatik.uni-leipzig.de

Abstract. The improvement of public health is one of the main indicators for societal progress. Statistical data for monitoring public health is highly relevant for a number of sectors, such as research (e.g. in the life sciences or economy), policy making, health care, pharmaceutical industry, insurances etc. Such data is meanwhile available even on a global scale, e.g. in the Global Health Observatory (GHO) of the United Nations's World Health Organization (WHO). GHO comprises more than 50 different datasets, it covers all 198 WHO member countries and is updated as more recent or revised data becomes available or when there are changes to the methodology being used. However, this data is only accessible via complex spreadsheets and, therefore, queries over the 50 different datasets as well as combinations with other datasets are very tedious and require a significant amount of manual work. By making the data available as RDF, we lower the barrier for data re-use and integration. In this article, we describe the conversion and publication process as well as use cases, which can be implemented using the GHO data.

Keywords: RDF, Health, WHO, GHO, statistics, datacube, CSV, spreadsheets

1. Introduction

The improvement of public health is one of the main indicators for societal progress. The *World Health Organization* (WHO)¹, a specialized agency of the United Nations, is mainly concerned with international public health with the main aim of the attainment of the highest possible level of health by all people. Besides publishing reports on global health problems, WHO also provides access to enormous amounts of statistical data and analyses for monitoring the global health situation. The WHO's *Global Health Observatory* (GHO)

publishes such statistical data and analyses for important health problems, which is categorised by either country, indicator or topic. The aim of GHO is to provide access to (1) country data and statistics with a focus on comparable estimates, and (2) WHO's analyses to monitor global, regional and country situation and trends².

GHO provides access to a wide variety of over 50 different datasets, such as the world health statistics, mortality and burden of disease, health expenditure per capita, deaths due to particular diseases such as HIV/AIDS, Tuberculosis, neglected tropical diseases, violence and injuries, health equity, just to name a few.

*Corresponding author. E-mail: zaveri@informatik.uni-leipzig.de.

¹<http://www.who.int/en/>

²<http://www.who.int/gho/about/en/>

Each dataset contains an extensive list of indicators which capture statistical data according to a region, country or based on gender. The data covers all the 198 WHO member countries³ and while some indicators are from the late 1970s onwards, some are prior to the mid-1990s. The data is updated as more recent or revised data becomes available or when there are changes to the methodology being used. A list of all the datasets with a description of its contents is provided in Table 1.

In this paper, we first describe the process of the conversion of the GHO data to RDF in Section 2. Details of the publishing and the interlinking GHO with other datasets are presented in Section 3. Section 4 portrays a few potential application scenarios and use cases for the GHO data. A number of related initiatives and how GHO is different than what already exists is discussed in Section 5. Finally, we conclude with the lessons learned in Section 6.

2. Dataset Conversion

The GHO data is published as spreadsheets describing a single data item (e.g. death, DALY) in several dimensions (e.g. country, population, disease). In order to convert the data to RDF, we used the RDF Data Cube Vocabulary [3] which is based on the popular SDMX standard⁴ and designed particularly to represent multidimensional statistical data using RDF. The vocabulary also uses the SDMX feature of content oriented guidelines (COG). COG defines a set of common statistical concepts and associated code lists that can be re-used across datasets.

However, transforming these spreadsheets to RDF in a fully automated way may cause information loss as there may be dimensions encoded in the heading or label of a sheet. Thus, we implemented a semi-automatic approach by integrating the algorithm as a plug-in extension in OntoWiki [1]. OntoWiki is a tool which supports agile, distributed knowledge engineering scenarios. Moreover, it provides ontology evolution functionality, which can be used to further transform the newly converted statistical data.

Using this plug-in⁵, when a spreadsheet containing multi-dimensional statistical data is imported into OntoWiki, it is presented as a table as shown in Figure 1.

Subsequently, the user has to manually configure the (1) dimensions, (2) attributes by creating them individually and selecting all elements belonging to a certain dimension and (3) the range of statistical items that are measured. Using RDFa, the corresponding COG concepts are automatically suggested, when a user enters a word in the text box provided. The specified configurations can also be saved as a template and reused for similar spreadsheets, such as for data published in consecutive years. Then the plug-in automatically transforms the data into RDF. A presentation detailing the conversion process is available⁶.

After converting the GHO data, an RDF dataset containing almost 8 million triples (number of triples for each individual dataset is reported in Table 1) was obtained and published at: <http://gho.aksw.org/>. The mortality and burden of disease dataset in GHO alone accounts for 3 million triples. An example of the death value 127 represented as RDF using the Data Cube vocabulary is illustrated in the following listing:

```

1 gho:Country    rdfs:subClassOf qb:DimensionProperty;
2               rdf:type rdfs:Class;
3               rdfs:label "Country" .
4
5 gho:Disease    rdfs:subClassOf qb:DimensionProperty;
6               rdf:type rdfs:Class;
7               rdfs:label "Disease" .
8
9 gho:Afghanistan  rdf:type ex:Country;
10                rdfs:label "Afghanistan" .
11
12 gho:Tuberculosis  rdf:type ex:Disease;
13                 rdfs:label "Tuberculosis" .
14
15 gho:c1-r6       rdf:type      qb:Observation;
16                 rdf:value    "127"^^xsd:integer;
17                 qb:dimension gho:Afghanistan;
18                 qb:dimension gho:Tuberculosis .

```

Listing 1: RDF representation of the death value '127' using the RDF Data Cube Vocabulary

3. Dataset Publishing and Linking

Dataset Publishing. After converting the GHO data as RDF, we published it as Linked Data using the *OntoWiki* platform [1]. OntoWiki not only allows the publishing and maintenance of the data but also provides a SPARQL endpoint for the dataset in combination with *Virtuoso*⁷ as the storage solution for the RDF model.

³<http://www.who.int/countries/en/>

⁴<http://sdmx.org>

⁵Available at aksw.org/Projects/Stats2RDF

⁶<http://goo.gl/OHDM9>

⁷<http://virtuoso.openlinksw.com/>

Table 1

Different statistical datasets available in the Global Health Observatory.

Dataset	Description	Triples #
Environmental health	Number of deaths due to children health, climate change, household air pollution, UV radiation, water, sanitation and hygiene	31,012
Epidemic prone diseases	Number of reported cases of cholera, meningococcal meningitis and statistics from the Global Influenza Surveillance and Response System	255,957
Equity	Equity figures for women health, urban health and social determinants of health	324,445
Health-related Millennium Development Goals	Health indicators associated with poverty and hunger, child mortality, maternal health, environment sustainability, and global partnership for development.	784,346
Health systems	Data on healthcare infrastructure, essential health technologies, aid effectiveness, health financing, essential medicines, service delivery and health workforce	234,340
HIV/AIDS	Data on the size of the epidemic and on the HIV/AIDS response	99,476
Immunization	Country and regional data of immunisation efforts for several diseases	625,082
Injuries and violence	Number of deaths due to road traffic accidents, data on demographic and socio-economic statistics, emergency care provision and existence of a national policy for human safety	242845
Mortality and burden of disease	Number of deaths, DALYs, life expectancy, mortality and morbidity, disease and injury country estimates for each country	3,000,000
Neglected Tropical Diseases	Statistics on newly reported cases of each of the neglected tropical disease that is monitored	167,841
Noncommunicable Diseases	Mortality measures, risk factors and health system response and capacity for each of the non-communicable disease that is monitored	1,409,629
Tobacco Control	Data on the prevalence of adult and youth consuming tobacco and various measures to help prevent tobacco consumption, such as policies, help, warnings, enforcing bans	379,283
Tuberculosis	Cases of incidence and mortality, diagnosis, drug regimens, treatment success for tuberculosis in each country	67,479

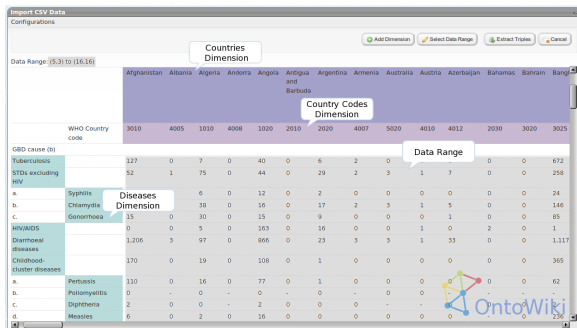


Fig. 1. Screenshot of the OntoWiki statistical data import wizard displaying a GHO table configured for conversion into RDF.

Additionally, it is also possible to browse the data with the HTML output of OntoWiki. Details and links of the SPARQL endpoint, the version, licensing, availability and link to the VoID file are listed in Table 2.

Table 2

Technical details of the GHO RDF dataset.

URL	http://gho.aksw.org/
Version date and number	01-11-2010, 1.0
Licensing	WHO allows reproduction of its data for non-commercial purposes.
VoID File	http://db0.aksw.org/downloads/void.ttl
The DataHub entry	http://thedatahub.org/dataset/gho

Dataset Interlinking. The URI's for diseases and countries are uniform for all the datasets in GHO since all the datasets utilize the same disease and country name in the original data. Thus, there was no explicit interlinking required within the datasets. A sin-

gle country or disease URI was therefore automatically linked to all the instances from the different tables associated with that country or disease and vice-a-versa. Thus, when a single country or diseases is looked up, all the corresponding instances from all the tables can be retrieved. There are a total of 192 unique country URIs and 116 disease URIs in the entire dataset.

We used the *mortality and burden of disease dataset* from GHO as a test-environment for link generation and linked it with the *LinkedCT*⁸ (the Linked Data version of ClinicalTrials.gov) and *PubMed*⁹ (converted to Linked Data by the Bio2RDF project) datasets for diseases and countries.

We used the *Silk 2.0* [4] tool, which is developed for discovering relationships between data items within different knowledge bases, that are available via SPARQL endpoints. Silk includes a declarative language for specifying (1) the types of RDF links that should be discovered and (2) the conditions which the data items must fulfil in order to be interlinked. We used the *Jaro distance* as string metric where applicable and two confidence value thresholds: (1) Links above 0.95 confidence were accepted and (2) links between 0.90 and 0.95 were saved to a separate file for manual inspection. The number of interlinks obtained for countries and diseases is displayed in Table 3 along with the precision for each set of links. The precision value was calculated manually by a researcher with a biomedical background by going through each links and evaluating the correctness. Then, using the formula: $(\text{correct links} * \text{total no. of links} / \text{total no. of links})$, the precision values were noted. Since a gold standard dataset was not available, the recall values were not calculated.

Additionally, we interlinked the diseases from GHO with the diseases in BioPortal¹⁰, in particular the ICD-10 codes¹¹. We used the *LIMES* [2] framework to create these links since the *SILK* tool did not provide the flexibility and efficiency that was required to generate these interlinks. There was a total of 107 interlinks that were created, of the total 116 diseases. Interlinks for diseases such as "Communicable, maternal, perinatal and nutritional conditions", "Maternal conditions" or "Nephritis and nephrosis" could not be found either because there was no match found or they were too generalized whereas ICD contained a thorough classi-

fication of the diseases. Moreover, in order to increase the interlinking between GHO and other datasets, we interlinked the countries from GHO with those in the DBpedia¹² and World Bank¹³ datasets. It is to be noted that for these three interlinks (ICD-10, DBpedia and World Bank), although LIMES was able to correctly find interlinks for the diseases and countries, there were some instances for which no links were found. For these the interlinks were manually created.

In addition to the ability to explore the data by using SPARQL and the resulting lists of resources, users are able to visualize the data by using *CubeViz*. CubeViz is an OntoWiki extension, which uses DataCube resources as input. After the selection of desired dimension properties such as `gho:disease` and `gho:country` as well as the measure property `gho:incidence` CubeViz is able to generate different type of charts (e.g. bar chart, pie chart, spline chart)). As an example, the incidence of the disease "Migraine" in selected countries can be visualized with CubeViz as depicted in Figure 2.

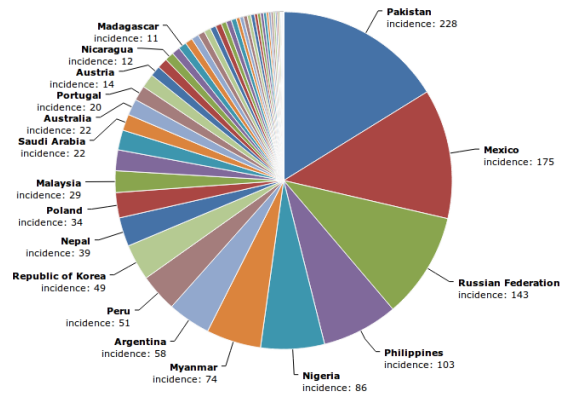


Fig. 2. Screenshot of CubeViz displaying the pie chart of incidence of Migraine in a subset of countries.

4. Application Scenarios and Use-Cases

In this section, we outline selected application scenarios and use-cases for the Linked GHO data.

⁸<http://linkedct.org/>

⁹<http://bio2rdf.org/>

¹⁰<http://www.bioontology.org/>

¹¹<http://www.who.int/classifications/icd/en/>

¹²<http://wiki.dbpedia.org/Downloads38#links-to-gho>

¹³<http://worldbank.270a.info/>

Table 3

Number of inter-links and precision values obtained between GHO and PubMed, LinkedCT, ICD-10, DBpedia and WorldBank for diseases and countries.

Links between	Link type	Source dataset	Source Instances	Target Dataset	Target Instances	Accepted Links	Verified Links	Precision Accepted	Precision Verified
Diseases	owl:sameAs	LinkedCT	5000	GHO	116	163	43	0.9625	-
Diseases	rdfs:subClassOf	LinkedCT	5000	GHO	116	469	45	1	0.9926
Diseases	owl:sameAs	GHO	116	PubMed	23618	453	75	1	0.7083
Diseases	owl:sameAs	GHO	116	ICD-10	4913	107	-	1	-
Locations	redd:locatedIn	LinkedCT	757341	GHO	192	300000	0	1	-
Countries	owl:sameAs	GHO	192	PubMed	23618	201	12	1	0.9583
Countries	owl:sameAs	GHO	192	DBpedia	2710	192	-	1	-
Countries	owl:sameAs	GHO	192	WorldBank	214	189	-	1	-

Monitoring health care scenarios. Since GHO provides information on mortality, morbidity, health status, service coverage and risk factors for several diseases in each country, it can be used by each country to monitor the disease prevalence for any given year and to compare prevalence as well as the effect of countermeasures with similar or neighbouring countries.

For example, Listing 2 shows the SPARQL query for retrieving the number of deaths due to Measles in all countries. Listing 3 shows the SPARQL query for retrieving the measles immunization coverage among 1-year-olds (%)¹⁴. The values retrieved from these SPARQL queries can be used to compare the number of deaths and the immunization services carried out in a country for a particular year to gauge its effectiveness.

```

1 SELECT ?deaths ?diseasename ?countryname
2 FROM <http://ghocountry.org>
3 FROM <http://interlinks.org>
4 WHERE {?item a qb:Observation.
5 ?item gho:Country ?country .
6 ?country rdfs:label ?countryname.
7 ?item gho:Disease ?disease.
8 ?disease rdfs:label ?diseasename.
9 ?item att:unitMeasure gho:Measure2 .
10 ?item eg:incidence ?deaths .
11 FILTER regex(?diseasename, "Measles")

```

Listing 2: SPARQL query for retrieving the number of deaths due to Measles in all countries.

This can help to implement either precautionary measures if the mortality is high or curb health expendi-

tures for diseases which seem to have adequate treatment options.

```

1 SELECT DISTINCT ?countryname ?incidence ?whichyear
2 WHERE {?item a qb:Observation.
3 ?item ex:incidence ?incidence.
4 ?item gho:Country ?country.
5 ?item gho:Year ?year.
6 ?year rdfs:label ?whichYear.
7 ?country rdfs:label ?countryname
8 FILTER regex(?whichYear, "2004")}

```

Listing 3: SPARQL query for retrieving the measles immunization coverage among 1-year-olds (%).

Disparity Analysis. Another application of the GHO dataset is evaluating the disparity between the availability of treatment options and the global burden of disease, as illustrated in the ReDD-Observatory project [5]. This project interlinks GHO with the PubMed and LinkedCT datasets so as to enable the evaluation of the disparity. This disparity is partially caused due to the limited access to information that would allow health care and research policy makers make more informed decisions regarding health care services. The hindrance lies in reliably obtaining and integrating data regarding the disease burden and the respective research investments. Therefore, as the Linked Data paradigm provides a simple mechanism for publishing and interlinking structured information on the Web, an opportunity is created to reduce this information gap that would allow for better policies in response to these disparities.

Listing 4 provides an example of a SPARQL query which retrieves the number of deaths (from GHO) and

¹⁴All prefixes can be found on <http://gho.aks.w.or> or on prefix.cc

the number of clinical trials (from LinkedCT) for the disease Tuberculosis and HIV/AIDS in all countries.

```

1 SELECT ?countryname ?diseasename ?value
2 count(?trial)
3 FROM <http://gho.aksw.org/>
4 FROM <http://linkedct.org/>
5 WHERE {?item a qb:Observation ;
6             gho:country ?country ;
7             gho:disease ?disease ;
8             att:unitMeasure gho:Measure ;
9             gho:incidence ?value .
10 ?country rdfs:label ?countryname .
11 ?disease rdfs:label ?diseasename .
12 ?trial a ct:trials ;
13        ct:condition ?condition ;
14        ct:location ?location .
15 ?condition owl:sameAs ?disease .
16 ?location shv:locatedIn ?country .
17 FILTER (?diseasename("Tuberculosis", "HIV/AIDS"))}

```

Listing 4: SPARQL query for retrieving the number of deaths and number of trials for Tuberculosis and HIV/AIDS in all countries.

Primary source providing ground truth. GHO enables direct linking to the ground truth data for secondary (e.g. scientific publications) or tertiary (e.g. encyclopedias) sources. This enables improved provenance tracking in those sources. It also allows automatic syndication of the data using SPARQL or simple REST queries, which enables a simpler verification of statements compared to the manual work, which would be necessary without Linked Data. For example, the Wikipedia entry for Disease¹⁵ (a tertiary source) uses the statistical values from GHO, in particular from the *mortality and burden of disease* data. The years of potential life lost (YPLL) and disability-adjusted life year (DALY) metrics are derived from GHO to provide information about them for several diseases categories in different regions of the world. Similarly, the corresponding DBpedia entry¹⁶ (also a tertiary source) also links to the GHO page about the burden of diseases¹⁷.

Human development data warehouse. Just as data warehouses and business intelligence are now integral parts of every larger enterprise, the linked GHO data can be the nucleus for a human development data warehouse. In such a human development data warehouse, a large number of statistical data and indicators are published by different organizations that could be

integrated automatically or semi-automatically in order to obtain a more interactive picture of the human development.

Currently, the indicators (e.g. the Human Development Index) are very coarse-grained, mainly referring to countries. Using linked data, such indicators could be computed on a much more fine-grained level, such as for cities and regions as well as with regard to different groups of people (e.g. per gender, ethnicity, education level). Policy making would be based on more rational, transparent and observable decisions as it is advocated by evidence-based policy.

For example, Listing 5 shows the SPARQL query for retrieving the public health expenditure (from the World Bank dataset) and the number of DALYs (Disability-Adjusted Life Years)¹⁸ caused by all diseases (from GHO) in the year 2004. The results from these queries can thus be compared per country per year to obtain an overview of the human development problems affecting each country.

```

1 SELECT ?countryNameGHO ?daly
2 (?obsValue AS ?publicHealthExpenditure)
3 WHERE {
4   GRAPH g-indicators: {
5     ?obs property:indicator indicator:SH.XPD.PUBL;
6     sdmx-dimension:refArea ?countryWB;
7     sdmx-dimension:refPeriod <http://
8       reference.data.gov.uk/id/year/2004>;
9     sdmx-measure:obsValue ?obsValue.
10  }
11  GRAPH g-meta: { ?countryWB a dbo:Country .}
12  SERVICE <http://gho.aksw.org/sparql> {
13    SELECT DISTINCT ?countryNameGHO ?daly {
14      ?item a qb:Observation.
15      ?item gho:Country ?countryGHO .
16      ?countryGHO owl:sameAs ?countryWB.
17      ?countryGHO rdfs:label ?countryNameGHO.
18      ?item gho:Disease ?disease.
19      ?disease rdfs:label "All Causes".
20      ?item att:unitMeasure gho:Measure2 .
21      ?item eg:incidence ?daly .
22    }
23  }
}

```

Listing 5: SPARQL query for retrieving the public health expenditure (from World Bank dataset) and number of DALYs caused by all disease (from GHO) in the year 2004.

5. Related Initiatives

There are already a number of efforts to convert health care and life science related data sets to Linked

¹⁵<http://en.wikipedia.org/wiki/Disease>

¹⁶<http://dbpedia.org/page/Disease>

¹⁷http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/index.html

¹⁸http://dbpedia.org/page/Disability-adjusted_life_year

Data such as LODD, LinkedCT, OBO ontologies and the W3C's Health Care and Life Sciences Working Group, each of which is discussed in this section along with the importance of converting and publishing the GHO datasets.

LODD, i.e. the Linking Open Drug Data project¹⁹, mainly converts, publishes and interlinks drug data that is available on the web, ranging from impacts of drugs on gene expression to results of the clinical trials. A number of datasets have been converted in this project²⁰ including DrugBank, DailyMed, SIDER to name a few. However, these datasets are restricted to drug data and even though they do contain disease data (from the Disease dataset), they do not connect the number of deaths or the health expenditure or the status of the health system in each country for each of the diseases that are included (as provided by GHO).

LinkedCT is the Linked Data version of Clinical-Trials.gov which publishes data about clinical trials in RDF and links it to other datasets such as PubMed. Even though, in LinkedCT each trial is associated with a disease and a drug, it does not provide information about the prevalence of the disease in a particular country, which is provided in GHO.

OBO is the Open Biological and Biomedical Ontologies project²¹ which aims to create a suite of interoperable reference ontologies in the biomedical domain. It brings together biology researchers and ontology developers who work together to develop a set of ontologies as well as design principles that can help develop interoperable ontologies. However, most of the ontologies developed are at the experimental level or organismal level and are not yet sufficiently interlinked with other datasets available as Linked Data. Additionally, the *NCBO's BioPortal*²² contains a large collection of controlled medical terminologies, all available as Linked Data.

*The Semantic Web Health Care and Life Sciences (HCLS Interest Group)*²³ is established by the World Wide Web Consortium (W3C) to support the use of Semantic Web technologies in health care, life sciences, clinical research and translational medicines. The group focuses on aiding decision-making in clinical research, applying the strengths of Semantic Web technologies to unify the collection of data for the

purpose of both primary care (electronic medical records) and clinical research (patient recruitment, study management, outcomes-based longitudinal analysis, etc.). Subgroups, on the other hand, focus on making the biomedical data available in RDF, dealing with biomedical ontologies, focus on drug safety and efficacy communication and support researchers in the navigation and annotation of the large amount of potentially relevant literature.

6. Conclusion and Lessons Learned

Although we were able to successfully convert the GHO dataset and utilize one of the datasets in a use case, we encountered some problems such as cumbersome conversion, low interlinking quality and lack of time series capability in the datasets. We discuss these problems in the sequel.

Conversion. The conversion process was cumbersome and time consuming because, first of all, each individual Excel files needed to be downloaded from the GHO web portal. Then, each file had to be converted into CSV so that it could be appropriately displayed as an HTML table in OntoWiki. Since the conversion method was semi-automated, one had to individually selected the dimensions, attributes and data range for each of the files. While some of the required steps such as the annotation of the CSV files for conversion are not automatable, other steps, such as the Excel to CSV conversion can be performed more efficiently (e.g. in a batch or through bulk processing).

Coherence. The number of interlinks obtained between the datasets for diseases was relatively low, as presented in Table 3. The main reason was the different use of identifiers for the naming of the disease. For example, 'heart attack' in GHO could not be matched with 'cardiac arrest' in LinkedCT using the basic string similar functionality of SILK. In order to address this problem, we plan to extend current linking tools in such a way, that background knowledge in the form of gazetteers can be also taken into account. Also, we plan to link the disease names with their corresponding ICD codes so as eliminate the need for entity recognition and also to improve precision and recall.

Temporal Comparability. The data in GHO is not published regularly every year. Also, since the health data recording and handling systems differ between countries, comparability of the data is limited. This is

¹⁹<http://www.w3.org/wiki/HCLSIG/LODD/>

²⁰<http://www.w3.org/wiki/HCLSIG/LODD/Data>

²¹<http://obofoundry.org/>

²²<http://bioportal.bioontology.org/>

²³<http://www.w3.org/blog/hcls/>

mainly due to the differences in definitions and/or time periods and incomplete data provision from different countries. Therefore, computing time trends is not possible using GHO, which would be a good indicator of the health scenario in each country over a number of years. We expect, however, that the increased visibility and transparency of a Linked Data version of GHO together with the enhanced possibility of annotation and linking (when compared to simple Excel sheets) will contribute to standardization and increased temporal comparability in the future.

Exploring GHO. Using OntoWiki or similar tools (such as Disco or Tabulator etc.) to browse the RDF data helps users to gain new insights. CubeViz as an OntoWiki extension provides visualization of statistical data (such as GHO) in an user friendly way by means of displaying the data in various types of diagrams and charts. However, a limitation of such generic visualization tools is their limited scalability.

Conclusion. In conclusion, by providing the GHO data as Linked Data and interlinking it with other datasets, it is possible to not only obtain information on important health related topics in each country but also ease the work of health care professionals for data analysis in providing easy access to data. Moreover, it provide opportunities to link to related data and thus perform analyses for current priority health issues. The Linked Data publishing and linking of the GHO data is a first milestone in a larger research and development agenda: The creation of a global human development data warehouse, which allows to interactively monitor social, societal and economic progress on a global scale.

7. Acknowledgment

This work was supported by grants from the European Union's 7th Framework Programme provided for the projects LOD2 (GA no. 257943) and LATC (GA no. 256975). We thank the LATC and LOD2 consortium members as well as the European Commission for their feedback on the GHO dataset.

References

- [1] S. Auer, S. Dietzold, and T. Riechert. OntoWiki – A Tool for Social, Semantic Collaboration. In I. Cruz, S. Decker, and D. Allemang, editors, *The Semantic Web – ISWC 2006, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, Proceedings*, volume 4273 of *Lecture Notes in Computer Science*, pages 736–749, Berlin / Heidelberg, 2006. Springer.
- [2] A.-C. Ngonga Ngomo and S. Auer. LIMES – A Time-Efficient Approach for large-scale link discovery on the Web of data. In T. Walsh, editor, *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 2312–2317. IJCAI/AAAI, 2011.
- [3] J. Tennison, R. Cyganiak, and D. Reynolds. The RDF Data Cube vocabulary. Technical report, W3C Working Draft 05 April, 2012. <http://www.w3.org/TR/vocab-data-cube/>.
- [4] J. Volz, C. Bizer, M. Gaedke, and G. Kobilarov. Discovering and Maintaining Links on the Web of Data. In A. Bernstein, D. Karger, T. Heath, L. Feigenbaum, D. Maynard, E. Motta, and K. Thirunarayan, editors, *ISWC 2009, Proceedings of the 8th International Semantic Web Conference, Chantilly, VA, USA, October 25-29, 2009*, volume 5823, pages 650–665, Berlin, Heidelberg, 2009. Springer-Verlag.
- [5] A. Zaveri, R. Pietrobon, S. Auer, J. Lehmann, M. Martin, and T. Ermilov. ReDD-Observatory: Using the web of data for evaluating the research-disease disparity. In J. F. Hübner, J.-M. Petit, and E. Suzuki, editors, *Proceedings of the 2011 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT 2011, Campus Scientifique de la Doua, Lyon, France, August 22-27, 2011*, volume 1, pages 178–185. IEEE Computer Society, 2011.