GI_Forum 2016, Vol.1
Page: 355-359
Short Paper
Corresponding Author:
frans.rip@wur.nl
DOI: 10.1553/giscience2016_01_s355

Some Thoughts on the Possible Benefits of Using VirLaBoK for GI Teaching Design and Locating GI Courses

Frans I. Rip
Wageningen University, The Netherlands

Abstract

This paper outlines the idea of storing descriptions of courses or curricula in the geographic information domain on the web as Linked Data. For this to become possible, each educational course offered should use standard terminology, as provided by a Body of Knowledge. This would facilitate course-finding on the web. The suitability of courses located for the needs of students could be assessed by data quality assessment from a fitness-for-use perspective.

Keywords:

geographic information, body of knowledge, course description, linked data, data quality

1 Introduction

Situation

Imagine a person seeking a training or education opportunity in the field of Geo-Information (GI). Gathering and comparing information about GI courses on offer is not an easy task because of the individual approaches of many teaching organizations in naming and describing their courses. In addition to that, the GI teaching subjects that people search for (Wallentin et al., 2014) do not always match up, in terms of terminology, with the education offers (Rip et al., 2014). Therefore, because of differences in the use of terminology, a web search for GI-learning opportunities will produce an incomplete collection of offers. Language differences also play a role. This situation does not support the ambition of the European Union for more international student mobility (EUPO, 2015) and cross-border trade (Euronews, 2012).

The idea

This paper outlines the idea for a GI Course Portal, based on a connection between the VirLaBoK software, Linked Data technology, and the data quality approach. Such a portal should make it easier to find GI-related courses or other learning opportunities, regardless of national boundaries and languages. It will then be up to organizations that offer such courses

to decide whether they want to be found by potential clients. The technology already exists, in the form of search engines. It is a matter of awareness, acceptance, and use of a knowledge standard for the GI domain, which should bring about greater harmony between the course descriptions of institutions offering GI teaching. Ultimately, all descriptions of GI learning opportunities should be accessible through one web portal.

Three components

The potential for VirLaBoK to deliver considerable improvement in the use of consistent terminology for GI teaching subjects looks promising. VirLaBoK is a software tool for designing GI courseware (Hossain et al., 2014), designed to work in conjunction with the American GI-BoK2 tools (Ahearn et al., 2013). Both products are based on the original Geographic Information Science & Technology Body of Knowledge (hereafter referred to as "GI-BoK"), originally published as a printed book (DiBiase et al., 2006). The improvement expected will be brought about by a wider use of standard terminology to describe GI teaching.

Another source of inspiration for this idea is Linked Data technology, especially the proposed five-star rating to express the level of linkage between data (Berners-Lee, 2010). This paper suggests that creating 5-star Linked Data about GI courses in Europe would benefit the seekers of GI courses by producing more results and 'tolerating' informal search terms in more languages. This would improve the visibility of organizations and the GI courses on offer.

The increased number of search results when looking for interesting courses on GI subjects will trigger the need for methodical assessment of their suitability. A useful approach might be that of Data Quality from a fitness-for-use perspective (Vullings et al., 2015).

2 More about the components

Virtual Lab for the BoK: VirLaBoK

This software tool is designed to compile the subject contents of a teaching unit (i.e. a training course or a curriculum). The tool was developed in the context of the Geographic Information: Need to Know project ("GI-N2K"), an EU-funded undertaking from 2013 to 2016 with partners from 25 European countries. The main pillars of VirLaBoK are an updated and 'Europeanized' version of the 2006 GI BoK, and a digitized American version of 2006 GI BoK called GI S&T BoK2 (Ahearn et al., 2013). The ontology-based content of BoK2 is called BoKOnto. The essence of BoKOnto is a dynamic network of semantic relationships between the concepts of GI-BoK: Knowledge Areas, Units, Topics and Learning Objectives. The interface for BoKOnto is called the Semantic Wiki, which will also be used in VirLaBoK (Hossain et al., 2014).

Linked Data star ratings

In the context of the Semantic Web, the linking of data is a key component. Put simply, the linking amounts to declaring the relation between a subject and an object in the form of a

"triple". For instance: Mary (subject) is a (predicate) person (object). This expresses that "Mary" belongs to the class called "person" by using the RDF data model for Linked Data.

Predicates and classes together are called the terms of a vocabulary. A vocabulary (LOV: Linked Open Vocabulary) is specific to a particular domain or industry. Because vocabularies contain defined terms, they provide the semantic glue that makes data meaningful in a context. Relationships can be written in a computer-readable way using one of the RDF file formats, which makes them suitable for the RDF database query language SPARQL (W3C, 2013).

Berners-Lee (2010) created a 5-star scheme to express the level of linkage. If a dataset is placed on the web in any form (e.g. as an image of a spreadsheet), its rating is 1 star, just for making it public. The file gets 3 stars when it is also machine-readable, in a non-proprietary file format. The fourth star could be assigned if the W3C standards RDF and SPARQL were also used. To be entitled to all 5 stars, the data should also be linked to other people's data.

Formulating the relationships between concepts in this way amounts to creating an ontology (Gruber, 1992: an ontology is a formal specification of a shared conceptualization), which is the purpose of re-engineering the GI S&T Body of Knowledge into BoKOnto (Ahearn et al., 2103), as well as the objective of the European GI-N2K project.

Data quality from a fitness-for-use perspective

In this era of open data, the question of whether an available dataset would be fit for the intended use is of growing importance, because open data were not originally produced to meet the specifications of an open-data end-user, as the end-user is unknown to the data producer.

Vullings et al. (2015) sketched a methodology for assessing the fitness-for-use of datasets. The approach requires the end-user to specify aspects of quality: properties the dataset needs in order for its use to be productive. Many quality aspects could be specified. Wang & Strong (1996) defined four categories of data quality (DQ): Intrinsic DQ, Contextual DQ, Representational DQ and Accessibility DQ. Each category comprises a number of different aspects. For example, the Intrinsic category includes the aspects of Believability, Accuracy, Objectivity and Reputation. The metadata of geographic datasets are also aspects of quality, but they were outside the scope of Wang & Strong.

For each aspect of quality, a measurable standard must be formulated. The standards should reflect the data needs of a specific use case. The degree of importance accorded to the various aspects may vary, which can be taken into account by using the MoSCoW method. This comes down to assigning prioritization weights: Must have, Should have, Could have or Would have (Kuhn, 2009).

Then, for each standard of each quality aspect, it should be determined whether the dataset complies with that standard. The results, each combined with the priority class of the quality aspect, could be combined into a single score for the dataset. A high score would indicate that a dataset better fits the intended use.

3 Courses on offer

Data Quality of courses offered

Harmonized descriptions on the web of GI courses or curricula provided by universities, colleges or other providers could be approached as a dataset. In the case of universities and colleges, the descriptions of the teaching on offer are probably directed primarily at, and shaped for, their own communities. However, if the descriptions are on the web they can also be regarded as open data for people outside those communities, similar to training offered by commercial organizations.

Is the data fit for use in the case of someone seeking training or education in some GI subject field? Following the data quality assessment methodology mentioned above, a person could systematically check how well the offer fits his/her needs. Among the quality aspects will surely be: the subject matter being taught, the teaching language, course fee, duration, teaching level, timetable, whether what is on offer is by distance learning or classroom learning, and, if classroom learning, the location. Some aspects, like the precise subject matter, language or fee, will have a greater influence on the weighted final score of each course on offer.

Linked Data of courses on offer

In the world of Linked Data, many sets of structured data already exist on stable websites that can be linked to. These data sets include, for instance, the GeoNames database. The Open University in the UK also maintains a website with access to datasets that relate to "the publications, qualifications, courses and Audio/Video material produced at the Open University, as well as the people involved in making them" (Open University, 2016). For a GI course portal, each teaching offer should be re-designed using VirLaBoK and other vocabularies.

4 Outlook

A seeker of GI education hopes that his/her search terms find a match among the teaching subjects on offer. How would the seeker know which search terms to use? This could be accomplished by creating a GI Course Portal with a facility to connect the user's search terms to BoK terminology. If the courses on offer were described using that terminology, a match might be found. A second requirement would be that there is a link between BoK terminology and up-to-date information about courses on offer in the whole of Europe. A particular search might result in finding that courses of interest are on offer in only two foreign countries, by three organizations, not in the first six months, and only one of them in a language the potential participant understands. The organizational context (university, college, private enterprise), the duration, fee and other properties will also be shown. This would be much like how today hotel accommodation or flights can be found (and booked) on the internet by specifying key criteria. The essential first step for this to become possible is wide acceptance of the GI Body of Knowledge, with VirLaBoK as its ambassador.

References

Ahearn, S.C., Icke, I., Datta, R., DeMers, M.N., Plewe, B., & Skupin, A. (2013) Re-engineering the GIS&T Body of Knowledge, International Journal of Geographical Information Science, 27:11, 2227-2245. http://dx.doi.org/10.1080/13658816.2013.802324 ; accessed Jan. 29 2016.

Berners-Lee, T, 2010: Linked Data. http://www.w3.org/DesignIssues/LinkedData.html ; accessed Jan. 29 2016.

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Taylor Luck, A., Plewe, B., Wentz, E., (Eds.), 2006. Geographic information Science & Technology Body of Knowledge. Association of American Geographers, Washington, DC, USA. http://www.gi-n2k.eu/?wpdmdl=448 ; accessed Jan. 29 2016.

EUPO European Union Publications Office, 2015: The EU and the Bologna Process – working together for change. Brochure.
http://ec.europa.eu/education/library/publications/2015/bologna-process-brochure_en.pdf ; accessed Jan. 29 2016.

Euronews, 2012: EU to simplify cross-border trade rules. Newspaper article.
http://www.euronews.com/2012/09/17/eu-to-simplify-cross-border-trade-rules ; accessed Jan. 29 2016.

GI-N2K. GEOGRAPHIC INFORMATION: NEED TO KNOW project website: http://www.gi-n2k.eu/ ; accessed Jan. 29 2016.

Gruber, T. 1992: What is an Ontology? http://www-ksl.stanford.edu/kst/what-is-an-ontology.html ; accessed Jan. 29 2016.

Hossain, M.I., Reinhardt, W., Casteleyn, S., Huerta, J. 2014. Development of an European version of the GI S&T BoK. Presentation at 9th European GIS in Education Seminar (EUGISES), 4 7 September 2014, Cork, Ireland.
http://eugises2014.eugises.eu/wp-content/uploads/2014/09/Hossain_etal_EUGISES_2014.pdf ; accessed Jan. 29 2016.

Kuhn, J. 2009: Decrypting the MoSCoW Analysis. DITY Weekly Newsletter Vol.5 nr.44. http://www.itsmsolutions.com/newsletters/DITYvol5iss44.pdf ; accessed Jan. 29 2016.

LOV. Linked Open Vocabularies website. http://lov.okfn.org/dataset/lov/about ; accessed Jan. 29 2016.

Rip, F.I., Lammeren, R. van, Bergsma, A., 2014. Analysis of the supply of geospatial education and training. Wageningen University, the Netherlands. http://www.gi-n2k.eu/?wpdmdl=456 ; accessed Jan. 29 2016.

The Open University website: http://data.open.ac.uk/ ; accessed Jan. 29 2016.

Vullings, W., Bulens, J., Rip, F.I., Boss, M., Meijer, M., Hazeu, G., Storm, M., 2015. Spatial Data Quality: What do you mean? 18th AGILE Conference on Geographic Information Science, 9-12 June 2015 Lisbon, Portugal. http://www.agile-online.org/Conference_Paper/cds/agile_2015/shortpapers/87/87_Paper_in_PDF.pdf ; accessed Jan. 29 2016.

Wallentin, G., Hofer, B., Traun, C., 2014. Analysis of the demand for geospatial education and training - Results of the GI-N2K Demand Survey. University of Salzburg, Austria. http://www.gi-n2k.eu/?wpdmdl=455 ; accessed Jan. 29 2016.

W3C, 2013: SPARQL Query Language for RDF. https://www.w3.org/TR/rdf-sparql-query/ ; accessed Jan. 29 2016.

Wang R.Y., Strong, D.M. (1996): Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 1996 Vol.12 nr.4, pp.5-34.