# OPEN DATA COURSE PROJECTS IN BUSINESS INTELLIGENCE

**Lili Aunimo[1], Raine Kauppinen[1] and Hami Kekkonen[2]**

*[1]Haaga-Helia University of Applied Sciences (FINLAND)*
*[2]City of Helsinki (FINLAND)*

## Abstract

This paper discusses the exploitation of open data in course projects for business information students. Open data has been used for several years in teaching business intelligence for second-year students of business information technology. Alternatives for using open data would be the use of teacher-created data, educational example data from a commercial software provider or proprietary data from a company. All of the aforementioned data types prove to be challenging when the goal is to offer students project-based learning with authentic and reuse permitting data. This paper provides a description of a model for a course on business intelligence that extensively uses open data. It presents an example of a successful course project implementing the model. Lastly, the experiences and feedback from all stakeholders that took part in the course implementation are presented.

Keywords: open data, business intelligence, course project, project-based learning, industry collaboration, higher education

## 1 INTRODUCTION

This paper discusses the exploitation of open data in course projects for undergraduate students of business information technology. Open data has been used for several years in teaching business intelligence for second-year students at Haaga-Helia University of Applied Sciences. This paper presents a model for integrating open data into teaching and presents one case in detail. In the course project presented, several sources of open data are combined to produce an innovative outcome.

Open data is public data that is freely available for anyone to use [1]. The research questions that this study attempts to answer are as follows:

1. How can open data and open APIs be used in bachelors' and master's level education in business information technology?
2. What kind of course model would be successful from the point of view of all stakeholders?
3. How does the use of open data and APIs enhance students' learning?

The motivation behind this study is that there is little research on how open data can be used in a pedagogical setting and how it affects learning. However, it is important to study these aspects because:

1. Knowing what open data is and how it can be produced and exploited is part of the data literacy skills that future business information technology professionals and citizens need [2].
2. Open data is a valuable source for educators as it is free, reliable, up-to-date and highly available. In contrast to commercially owned educational data sources it does not suddenly disappear.
3. Open data is valuable for students because it is not limited by restrictive copyrights, but it can be freely used also beyond the course project.

The empirical part of the research is based on experiences gained while implementing the "Business Intelligence" course using open data for several years. Alternatives for using open data would be the use of teacher-created data, educational example data that typically comes with the software or that is available from commercial software providers such as Microsoft or proprietary data from a company. All of the aforementioned data types provide some challenges when the goal is to offer students project-based learning with authentic and reuse permitting data.

In addition to giving the students access to authentic data permitting reuse, the students of the course implementation of fall 2019 were offered close collaboration with the City of Helsinki. The City of Helsinki provides the data and coordinates the open data service, called Helsinki Region Infoshare (HRI). The students participated in the process of opening a new data set to the public by testing it and giving feedback about it. As the grande finale of their course project, the best student team was given the opportunity of presenting their project work at the Helsinki Loves Developers meetup organized by HRI.

The paper will provide a description of a model for a course on business intelligence that extensively uses open data. It will present a concrete example of a successful course project that combines several sources of open data into one meaningful project. Lastly, the experiences and feedback from all stakeholders that took part in the course implementation are presented.

## 2 RELATED WORK

Open data is public data that is freely available for anyone to use [1]. The main factors contributing to data openness are technical accessibility, free access, licensing permitting reuse, findability and understandability [3][4]. Typical sources for open data are public governance and science [5][6].

Data portals are the easiest way to access open data. In Finland, examples of open data portals include opendata.fi maintained by Digital and Population Data Services Agency of Finland [7] and HRI hri.fi maintained by the City of Helsinki [8]. The data may be offered from the portal, for example, as simple files in CSV (comma separated values) format or through APIs (application programming interface) [5, 212-213].

The state of the open data maturity in different countries can be assessed using open data indices such as Global Open Data Index [9] and Open Data Barometer [10]. These evaluate different open data portals from multiple viewpoints and they report open data maturity per societal sector or category such as education, public procurement or national map data. The different viewpoints mean factors such as is the data available without registering to the service as a user or is the data in a machine readable format. The Global Open Data Index uses 11 such viewpoints to evaluate the data openness in each category. In addition, the organizations behind the Open data Barometer and the Global Open data Index advocate local open data and open policies in public governance.

Open data and open APIs suit well courses that are taught using both the problem-based learning (PBL) and the project-based learning methodologies. PBL is a learning method which centers on the student and develops his active learning and problem-solving skills as well as his knowledge on the subject matter [11]. Project-based learning is an individual or group activity that goes on over a period of time, resulting in a product, presentation, or performance. It typically has a time line and milestones, and other aspects of formative evaluation as the project proceeds [12]. The main differences between problem and project-based learning are as follows [13]:

1. Project-based learning typically has a longer time span than PBL as the projects typically last for several months.
2. Project-based learning is typically carried out in the end of the studies in higher education because it resembles real professional projects and because it builds on a solid background in skills and knowledge of the field in question.
3. PBL has more focus on knowledge acquisition.

The scientific literature reporting the use of open data in education is relatively scarce. There have been some learning experiences in the fields of data journalism [14], teacher education [2] and science education in high school level [16]. Guy et al. [16] describe the activities carried out by the LinkedUp project aiming at the promotion of open data in education. Overall, three possible approaches in using open data in education have been identified [14, 153] [16]:

1. To use open data and the connected technologies to make educational tools more efficient and flexible.
2. To publish learning resources as open data.

3. To use introduce the open data into the educational environment in order to provide students with more options to organize their schedule and learning process. For example, publishing university study module description and schedules as open data.
4. To use open data as both the learning topic and the material on which the educational process is actually built.

This paper and the examples provided focus on the fourth use of open data in education. A concrete example of the use of open data in higher education has been used is from a bachelor level course with educational sciences students using project-based learning was used [2]. The course had a unit requiring the use of open data portal, exploring open data visualizations and the raw open data, presenting the selected data and reflecting its value for a specific goal [2, 183-184]. Here, the use of open data can be seen as a way of enhancing learning related to data literacy that is an essential skill in modern society [2, 186].

In this case, the students expressed only moderate early interest in open data, possible because of a belief that data related activities are far from their professional identity in construction. However, based on the results, most of the students were able to use and extract useful visualizations although most students required support from the teacher to understand the concept of dynamic visualization. Also, the students understood the concept of open data as well as how original representations could be created from the data. A few were also able to create new visualizations extracted from raw data that was in CSV format. [2, 185-187]

The data journalism case from higher education reports the use of open data published in an open data portal by the Russian government. One part of this experiment is the use of open learning materials that help in using the data[11]. The third case explaining the use of open data in education explains how open data is gathered through an IoT device named senseBox [16] containing several sensors. This sensor data is then published as open data into a global web portal as part of a high school science project [16].

# 3 EXPERIMENTAL SETTING: CASE HRI AND OPEN VEHICLE SPEED DATA

## 3.1 The BI course project on open vehicle speed data

The bachelor level course “Business Intelligence” (BI) for students of business information technology has been implemented using open data for three years now. Alternatives for using open data would be the use of teacher-created data, educational example data from a software provider such as Microsoft or proprietary data from a company. However, each of the aforementioned data types provide problems when the goal is to offer students project-based learning with authentic and data permitting wide scale reuse.

In addition to giving the students access to authentic data permitting reuse, the students of the course implementation of fall 2019 were offered close collaboration with the City of Helsinki. The City of Helsinki provides the data and coordinates an open data service called HRI. The students participated in the process of opening a new data set to the public by testing it and giving feedback about it. As the grande finale of their course project, the best student team was given the opportunity of presenting their project work at the Helsinki Loves Developers meetup organized by HRI.

The cooperation process, stakeholders and the added value brought in by each stakeholder of the fall 2019 implementation is depicted in Figure 1. The data set was a vechicle speed data that the City of Helsinki was planning to publish as open data via hri.fi. The vehicle speed data is collected in several measurement points and the City of Helsinki was still figuring out the best format for publishing the data because of its huge volume and the massive processing needed.

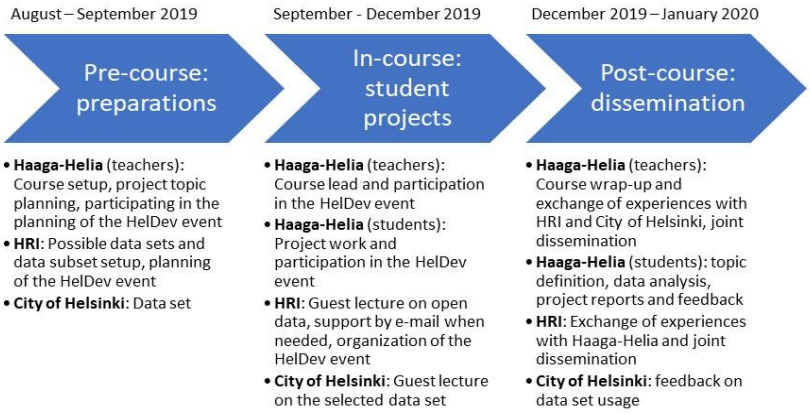


*Figure 1. The cooperation process phases, stakeholders and the value provided by each of the stakeholders.*

In the course project from fall 2019, a subset of the available vehicle speed data from the city of Helsinki was preselected to the students. It consisted of data from three selected measuring points and from specific date ranges for each of them.

The data is in CSV format with each measurement on its own row. The data from each measuring point was delivered as a separate file. The accuracy of the time is presented in seconds and the speed in whole numbers. In the case of several vechicles driving in a line, only the speed of the first vehicle is recorded. Vehicle types are not recorded. The total number of rows in the dataset is 4 212 529.

In the course projects, the task of the students was to both analyze the vehicle speed data and to join it with at least one other open dataset. Examples of other open data sources used by the students include weather data from the Finnish Meteorological Institute, traffic accident data, cyclist data, city bike data, passenger data from Helsinki Region Public Transport, and traffic volumes in Helsinki. In addition to these, also the MapSite from the National Land Survey of Finland. In the following are some student project topics:

1. Is there a relation between vehicle speed and weather conditions such as temperature, rain and visibility?
2. Is there a relation between vehicle speed and the amount of cyclists?
3. Is there a relation between traffic accidents and vehicle speed? Special attention was paid to the proximity of schools and to the times when pupils enter or leave the school premises.
4. How do special days such as midsummer or the Finnish national day as well as holiday seasons, weekends etc. affect vehicle speed?

## 3.2 HRI and university projects

HRI is the open data service of the cities of Helsinki, Espoo, Vantaa and Kauniainen. HRI was launched in 2011. The service provides data regarding the cities of the Helsinki Metropolitan Area and Helsinki region as open data for anyone to use freely. The service offers an abundance of, for example, geographic data and statistics through open application programming interfaces (APIs) or as

files. HRI service helps data owners to open their datasets, supports data users, maintains hri.fi website and data catalogue and arranges Helsinki Loves Developers meetups.

HRI service helps the cities open their public data but also encourages people to utilize open data, for example, in educational institutes and hackathons. Open data has already enhanced democracy, the creation of new business activities and intensified the city's own operations as well.

One of HRI's key target groups is students because they are usually very open-minded, think out-of-the box and utilize open data in innovative ways. It is important that students are informed about the potential of open data before they go into working life. HRI has successfully cooperated with different educational institutions for several years. The major ways of cooperating with universities include:

- Guest lectures for students and teachers on what is open data and how it can be used by citizens, businesses and public organizations.
- Course projects that typically last one semester. One was described in this paper.
- Offering help for students writing their thesis work. Open data is a good data source because it is reliable, local and up to date. The personnel have also given interviews to students regarding open data or provided them help with any other open data related issue.

# 4 RESULTS

## 4.1 A Model for a course design that combines several sources of open data

Based on the experiences of teaching BI using open data, a model for course design was developed. The model is depicted in Figure 2. The figure shows that planning a new course implementation begins with the careful study of the learning objectives. These are updated every time based on new requirements coming from the job market and based on the learning outcomes of the previous iteration as well as based on the new developments in the field under study. When the learning requirements are clear they are classified into those that have to be taught, either using problem-based learning or traditional lecturing. Our experiences tell that when the main target group is second year students, not all of the learning objectives can be met if only project based learning is applied. Open data is used both in the upper rectangle with relatively small and concrete learning objectives and in the lower rectangle that consists of larger and more complex learning objectives that are fulfilled by the project. Finally the learning outcomes of both pedagogical approaches are assessed and feedback is collected from the students, teachers and other possible stakeholders such as organizations producing open data or organizations willing to exploit it. This feedback is used when the planning of the next implementation starts.

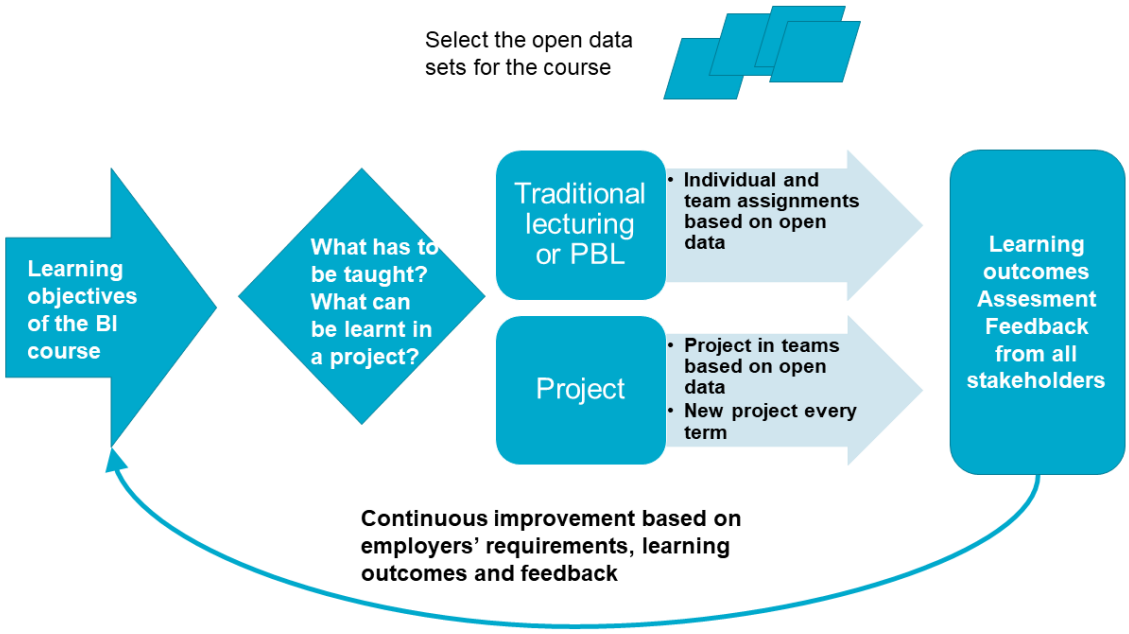


*Figure 2: The model for a BI course utilizing open data.*

## 4.2 The experiences from the BI course project utilizing open vehicle speed data

The experiences and feedback from all stakeholders that took part in the course implementation involving close collaboration with three parties: the students, the teachers and the organization that is in the process of opening its data and coordinates the open data service

The course with student projects based on a previously unexplored dataset was a very good experience for both HRI and Haaga-Helia. From the point of view of HRI, the feedback from the students and teachers regarding the dataset to be opened was very useful and it was forwarded to the data owner. From the point of view of both Haaga-Helia and HRI this cooperation is definitively worth continuing.

Experiences on using open data in the student projects were mainly positive. Especially the use of previously unpublished real-life data that was to be opened was interesting for both the students and the teachers. The feedback on data usage from the students concerned data uniformity, metadata, data formats and geographical coordinates. The uniformity of the available data was seen as an essential characteristic and this was mainly the case with the vehicle data. However, the vehicle data had varying date ranges in different measurement points. In addition, other open data sets used in the student projects had its own date ranges that did not always match the date ranges in vehicle data.

Another essential characteristic making open data feasible to use from the students' viewpoint was metadata describing and related to the actual data. This is important especially when the user of the data is not an expert in the domain the data is from as was the case with the students in their projects. Also, additional information on the way information has been collected and possible missing information or known errors would have been useful to the students.

The students were able to handle data in both CSV and JSON (JavaScript object notation) formats. The BI course does not require the use of APIs, the CSV format was used when available. In a case where the needed data was not available as CSV, it was noted that the loading and handling of the full data in JSON required quite a lot of work. It would have been good to be able to load only a part of it instead the full set.

Regarding the geographical coordinates, both the students and teachers felt that the coordinates used should preferably be available (also) in an international format such as WGS-84, since the tools used such as Microsoft Power BI and KNIME support those directly. While it is possible to transform the coordinates, this may slow down the application of open data when the potential users are not focused

on geographical data handling and transformations that is the case with the business information technology students.

The feedback from the students on the BI course was also mainly positive. It reflected the motivation gained from using the real-life data as well as the overall orientation towards working life including the co-operation with the HRI, project work and using the currently widely used BI tools. Regarding the project work, there was a suggestion to have a more uniform starting point and first phases for the projects even if it would meand more similar projects contentwise.

The BI course ended with a presentation in the Helsinki Loves Developers event where a teacher and a student shared the experiences on using the vehicle speed data and other open data in student projects. This event gained very good feedback from the praticipating students further widening their understanding on the availability and the possibilities of the usage of open data.

## 5 CONCLUSIONS

This work discusses the use of open data in education – especially the case where open data is one of the learning objectives of the courses and where open data is utilized in the course as study material.

The paper presented a model for courses using open data in teaching Business Intelligence to undergraduate students in business information technology. In addition to the model, one course testing the model and using several open data sets was presented. The experiences gathered from the course implementation were presented and discussed.

In the future the created model could also be implemented in master's level courses on business analytics in Haaga-Helia. International financial open data as well as open data sets from the World Bank have already been used in single learning assignments, but the course could benefit from a more wide and systematic use of open data.

## REFERENCES


[1] Open Knowledge Foundation, "The Open Definition", 2020. Retrieved from https://opendefinition.org/ (5.5.2020).

[2] J.E. Raffaghelli, "Open Data for Learning: A Case Study in Higher Education", *Exploring the Micro, Meso and Macro. Proceedings of the European Distance and E-Learning Network 2018 Annual Conference Genova*, pp. 178-190, 2018.

[3] R. Kitchin, *The Data Revolution*. London: Sage, 2014.

[4] S.R. Auer, C. Bizer, C., G. Kobilarov, J. Lehmann, R. Cyganiak and Z. Ives, "DBpedia: A Nucleus for a Web of Open Data". *The Semantic Web. ISWC 2007, ASWC 2007. Lecture Notes in Computer Science*, vol. 4825, pp. 722-735, 2007.

[5] N. Verma, M.P. Gupta, "Open Government Data: More than Eighty Formats", *Towards E-Governance in the Cloud. Frameworks, Technologies and Best Practices. 9th International Conference on E-Governance. ICEG 2012*, pp. 207-217, 2012.

[6] J.C. Molloy, "The Open Knowledge Foundation: Open Data Means Better Science", *PLoS Biology*, vol. 9, no. 12, 2011. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3232214/ (5.5.2020).

[7] Digital and Population Data Services Agency of Finland, "Avoindata.fi. All the Finnish Data from One Place", 2020. Retrieved from https://www.avoindata.fi/en (5.5.2020).

[8] Helsinki Region Infoshare, "Open Data Service. Making Better Use of Public Data in Helsinki Region", 2020. Retrieved from https://hri.fi/en_gb/ (5.5.2020).

[9] Open Knowledge Foundation, "Global Open Data Index", 2020. Retrieved from https://index.okfn.org/ (5.5.2020).

[10] World Wide Web Foundation, "Open Data Barometer", 2020. Retrieved from https://opendatabarometer.org/ (5.5.2020).

[11] Tandogan RO, Orhan A. The Effects of Problem-Based Active Learning in Science Education on Students' Academic Achievement, Attitude and Concept Learning. Online Submission. 2007;3(1):71-81

[12] Donnelly, R. Fitzmaurice, M. (2005) Collaborative Project-based Learning and Problem-based Learning in Higher Education: a Consideration of Tutor and Student Role in Learner-Focused Strategies. In G. O'Neill, S. Moore & B. McMullin (eds) Emerging Issues in the Practice of University Learning and Teaching (pp.87-98). Dublin, AISHE/HEA.

[13] Perrenet JC, Bouhuijs PA, Smits JG. The suitability of problem-based learning for engineering education: theory and practice. Teaching in higher education. 2000 Jul 1;5(3):345-58.

[14] I. Radchenko, A. Sakoyan, “On Some Russian Educational Projects in Open Data and Data Journalism”, *Open Data for Education. Lecture Notes in Computer Science*, vol. 9500, pp. 153-165, 2016.

[15] Thapa L, Naseer H, El-Kaiy S, Bartoschek T. Measuring Growth Conditions of Salad Plants Using Sensors: a High School Project. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences. 2018 Nov 20.

[16] Marieke Guy, Mathieu d’Aquin, Stefan Dietze, Hendrik Drachsler, Eelco Herder and Elisabetta Parod (2014) Linking open data for education. issn:1361-3200. http://www.ariadne.ac.uk/issue/72/guy-et-al/ [last accessed 7.5.2020]